DOCUMENT RESUME

ED 451 718 FL 026 657

AUTHOR	Fotos, Sandra, Ed.
TITLE	JALT Journal, 1999.
INSTITUTION	Japan Association for Language Teaching, Tokyo.
ISSN	ISSN-0287-2420
PUB DATE	1999-05-00
NOTE	159p.; Number 2 of Volume 21 was not issued in 1999.
AVAILABLE FROM	JALT Central Office, Urban Edge Building, 5f, 1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan (cover price: 950 yen). Tel: 03-3837-1630; Fax: 03-3837-1631; Web site: http://www.langue.hyper.chubu.ac.jp/jalt/pub/tlt.
PUB TYPE	Collected Works - Serials (022)
LANGUAGE	Japanese, English
JOURNAL CIT	JALT Journal; v21 n1 May 1999
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Adult Education; Class Activities; *Communicative Competence (Languages); *Computer Uses in Education; Discourse Analysis; Elementary Secondary Education; English (Second Language); Foreign Countries; Higher Education; *Pragmatics; Professional Development; Second Language Instruction; Second Language Learning; Self Evaluation (Individuals); *Student Adjustment; Student Evaluation; *Student Placement; Teaching Methods; World Wide Web
IDENTIFIERS	Japan

ABSTRACT

This journal (usually published twice a year) is a publication of the Japan Association for Language Teaching (JALT), a nonprofit professional organization of language teachers dedicated to the improvement of language learning and teaching in Japan. JALT's publications and events serve as vehicles for the exchange of new ideas and techniques, and a means of keeping abreast of new developments in a rapidly changing field. Each issue includes several sections and departments: feature articles, point to point articles where the major issues of the field are debated, research forum, perspectives, book reviews, and JALT journal information. Topics highlighted in this volume include the following: placement testing, measuring pragmatic competence, massive input, influences on intercultural adjustment, learner self assessment, LANs (local area network computers) and discourse quality, grades and self-efficacy, and effects of entrance examinations. (KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Gene von Troyer

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

SSN 0287-2420 ¥950

ED 451 718

全国語学教育学会

JALT JOURNAL

May 1999

Volume 21 • No. 1

Japan Association for Language Teaching

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Inside this issue:

Placement testing • Measuring pragmatic competence
Massive input • Influences on intercultural adjustment
Learner self-assessment
LANs and discourse quality • Grades and self-efficacy
Effects of entrance examinations

2

JALT Journal

Volume 21, No. 1
May, 1999

Editor
Sandra Fotos
Senshu University

Associate Editor
Nicholas O. Jungheim
Aoyama Gakuin University

Reviews Editor
Patrick Rosenkjar
Aoyama Gakuin University

Japanese-Language Editor
Shinji Kimura
Kwansei Gakuin University

Editorial Board

Charles Adamson
Miyagi University
Fred Anderson
Fukuoka University of Education
William Bradley
Ryukoku University
James Dean Brown
University of Hawai'i, Manoa
Christine Casanave
Keio University, SFC
Ann Ediger
Hunter College, City University of New York
Rod Ellis
University of Auckland
John Fanselow
Teachers College, Columbia University
Greta Gorsuch
Mejiro University
Dale T. Griffee
Seigakuin University
Paul Gruba
University of Melbourne
Thomas Hardy
Tamagawa University
Eli Hinkel
Seattle University
Guy Modica
Seikei University
Tim Murphey
Nanzan University
Mary Goebel Noguchi
Ritsumeikan University
Akito Ozaki
Nagoya University
Tim Riney
International Christian University
Carol Rinnert
Hiroshima City University
Peter Robinson
Aoyama Gakuin University
Tadashi Sakamoto
Nanzan University
Bernard Susser
Doshisha Women's Junior College
Satoshi Toki
Osaka University

Additional Readers: William Acton, Charles Browne, Steve Cornwell, Mary Flaherty, Noriko Iwashita, Yuri Kite, Virginia Lo Castro, Ethel Ogane, Tim Murphey, Tamara Swenson, Brad Visgatis

***JALT Journal* Proofreading**: Carolyn Ashizawa, Andrew Moody, Jack Yohay

JALT Publications Board Chair: William Acton

***The Language Teacher* Editor**: William Lee

Layout: The Word Works **Cover**: Amy Johnson **Printing**: Koshinsha

***JALT Journal* on the Internet**: http://www.als.aoyama.ac.jp/jjweb/jj_index.html

Contents

May 1999
Volume 21 • No. 1

3 In This Issue
4 From the Editors

Articles

7 Using a Commercially Produced Proficiency Test in a One-Year Core EFL Curriculum in Japan for Placement Purposes
Brent Culligan & Greta Gorsuch
29 Evaluating Six Measures of EFL Learners' Pragmatic Competence
Ken Enochs & Sonia Yoshitake-Strain
51 Massive Input Through Eiga Shosetsu: A Pilot Study with Japanese Learners
Michael "Rube" Redfield
66 Influence of Personality, L2 Proficiency and Attitudes on Japanese Adolescents' Intercultural Adjustment – *Tomoko Yashima*

Research Forum

87 Evaluating Learner Self-Assessment – *Colin Painter*

Perspectives

103 Raising the Quality of Discourse Using Local Area Networks in Returnee Classes – *John Herbert*
112 The Relationship between Self-Efficacy and Language Learners' Grades – *Stephen A. Templin*
125 A Myth of Influence: Japanese University Entrance Exams and Their Effect on Junior and Senior High School Reading Pedagogy
Bern Mulvey

Reviews

143 The Language Instinct: How the Mind Creates Language (Steven Pinker) – *Robert Blaisdell*
144 Testing in Language Programs (James Dean Brown) – *Ian G. Gleadall*
147 Using Corpora for Language Research (Jenny Thomas & Mick Short [Eds.]) – *Jim Ronald*
150 Teacher Cognition in Language Teaching: Beliefs, Decision-Making and Classroom Practice (Devon Woods) – *Kazuyoshi Sato & Tim Murphey*

JALT Journal Information

153 Information for Contributors

All materials in this publication are copyright ©1999 by their respective authors.

Japan Association for Language Teaching

The Japan Association for Language Teaching (JALT) is a professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and a means of keeping informed about new developments in the rapidly changing field of second and foreign language education. Established in 1976, JALT serves an international membership of more than 3,400, and there are 38 JALT chapters and one affiliate throughout Japan. JALT is the Japan affiliate of International TESOL (Teaching English to Speakers of Other Languages) and is a branch of IATEFL (International Association of Teachers of English as a Foreign Language).

JALT publishes *JALT Journal*, a semi-annual research journal, *The Language Teacher*, a monthly magazine containing articles, teaching activities, reviews and announcements about professional concerns, *JALT Applied Materials*, a monograph series, and *JALT International Conference Proceedings*.

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually. Local meetings are held by each JALT chapter and JALT's 13 Special Interest Groups (SIGs) provide information on specific concerns. JALT also sponsors special events such as workshops and conferences on specific themes, and awards annual grants for research projects related to language teaching and learning.

Membership is open to those interested in language education and includes enrollment in the nearest chapter, copies of JALT publications and reduced admission to JALT-sponsored events. For information, contact the JALT Central Office.

JALT National Officers, 1999

President: Gene van Troyer

Vice President: Brendan Lyons
Treasurer: David McMurray
Recording Secretary: Thomas Simmons

Program Chair: Joyce Cunningham
Membership Chair: Richard Marshall
Public Relations Chair: Mark Zeid

Chapters

Akita, Chiba, Fukui, Fukuoka, Gunma, Hamamatsu, Himeji, Hiroshima, Hokkaido, Ibaraki, Iwate, Kagawa, Kagoshima, Kanazawa, Kitakyushu, Kobe, Kumamoto (affiliate), Kyoto, Matsuyama, Miyazaki, Nagasaki, Nagoya, Nara, Niigata, Okayama, Okinawa, Omiya, Osaka, Sendai, Shinshu, Shizuoka, Tochigi, Tokushima, Tokyo, Toyohashi, West Tokyo, Yamagata, Yamaguchi, Yokohama.

Special Interest Groups

Bilingualism; Computer Assisted Language Learning; College and University Educators; Foreign Language Literacy (Affiliate SIG); Global Issues in Language Education; Japanese as a Second Language; Junior and Senior High School Teaching; Learner Development; Material Writers; Other Language Educators (Affiliate SIG); Professionalism; Administration and Leadership in Education; Teacher Education; Teaching Children; Testing and Evaluation; Video; Gender Awareness in Language Education (Forming SIG)

JALT Central Office

Urban Edge Building, 5F 1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel: 03-3837-1630; Fax: 03-3837-1631

In This Issue

Articles

This section has four articles. In the first paper **Brent Culligan** and **Greta Gorsuch** use analysis of item facility, item discrimination, and item difference indices to evaluate use of the SLEP test for placement purposes in a Japanese university EFL program. On the basis of their results, they make suggestions for modifications and supplemental procedures to produce a better "fit." Using Japanese university EFL learners, **Ken Enochs** and **Sonia Yoshitake-Strain** analyze the reliability, validity, and practicality of the multi-test framework measuring cross-cultural pragmatic competence developed at the University of Hawaii. They suggest that the tests are generally reliable and valid and are able to identify learners with extended overseas experience. In the next paper **Michael "Rube" Redfield** presents a pilot study using movie viewing and extensive reading of "*Eiga shosetsu*," movie tie-in novels, to provide massive comprehensible input for Japanese university EFL learners. The learners who participated in the project made significant gains on reading, listening and vocabulary identification measures. In the last paper **Tomoko Yashima** explores the influence of target language proficiency and extroversion on the intercultural adjustment process of Japanese high school sojourners in the United States. She finds that extroversion predicts student self-measures of adjustment, whereas English proficiency predicts adjustment as rated by the students' host families.

Research Focus

In this section, **Colin Painter** reports the results of an exploratory correlational analysis of student self-assessed scores compared with teacher scores, suggesting that the significant correlations observed indicate the reliability of the self-assessment process.

Perspectives

Examining use of a local area network (LAN) in a "returnee" class at a Japanese university, **John Herbert** finds that classroom discourse is enhanced since students can work at their own pace and participate more freely online than in regular oral activities. **Stephen Templin** uses questionnaire analysis to examine whether Japanese EFL learners with high self-efficacy perform better in class than students with a lower belief in their abilities to accomplish language tasks. In the final paper **Bern Mulvey** uses the results of analysis of the research literature to challenge the idea that entrance examination "washback" determines Japanese high school foreign language reading pedagogy and textbook content.

Reviews

Topics covered in book reviews by **Robert Blaisdell**, **Ian Gleadall**, **Jim Ronald**, and **Kazuyoshi Sato** and **Tim Murphey** include the cognitive origins of language, testing in language programs, the use of language corpora, and the relationships of teacher beliefs, assumptions and knowledge with teaching practice.

From the Editors

With this issue **Patrick Rosenkjar** takes over as Reviews Editor and former Reviews Editor **Thomas Hardy** joins the Editorial Advisory Board. We also welcome new Editorial Board member **Tim Murphey** and new proofreaders **Carolyn Ashizawa** and **Andrew Moody**.

Conference News

The 25th JALT Annual Conference on Language Teaching/Learning and Educational Materials Exposition will be held October 8-11, 1999, at the Maebashi Green Dome, Maebashi-shi, Gunma-ken. The Conference theme is "Teacher Action, Teacher Belief: Connecting Research and the Classroom." Contact the JALT Central Office for information.

Corrections

Part of a sentence in author Ron Grove's book review in Vol. 20 (1), p. 128-9, was omitted. The sentence should read:

> Just as it would be impossible to discuss pronunciation without concepts like "voiced/unvoiced" or "stop/continuant," it was necessary for Brazil to develop terminology appropriate for discussion of intonation, and this may be his most lasting contribution.

The title of the Japanese-language article by Shinichiro Yokomizo in Vol. 20 (1), pp. 37-46, was given incorrectly in the text. The correct title should read:

「発話矯正：コミュニティ・ランゲージ・ラーニングの理論の応用」

In addition, Mr. Yokomizo's biodata and Table 4 were omitted. We sincerely apologize for any inconvenience this has caused and print them below.

ハワイ大学マノア校で言語学(日本語)MA及びPh.D. を取得。1986年より同校で日本語を教え、現在は南山大学外国人留学生別科講師。CLL・ドリルのコンテクスト・アクションリサーチを専門とする。

図4　各教室活動における発話矯正のガイドライン（○=使うべきである，△=使ってもよい，×=使うべきではない）

			機械的ドリル		ペア・ワーク			ロール・プレイ	
教育目標			正確さ		なめらかさ			なめらかさ	正確さ
					学習者同士の・矯正	自信のない学習者	自信はあるが、確認を望む学習者	対話中	対話終了後
CLLの5段階			1または2		3	2	4	2または4	4
誰が誤りを直すべきか	教師		○		学習者が決定する	○	○	○	△
	誤りを犯した学習者本人		△			△	△	△	○
	他の学習者		△			△	△	○	○
どの誤りを直すべきか	誤り	全体的誤り	○			○	○	×	○
		局部的誤り	○			○	○	×	○
		否定的な評価につながる誤り	○			○	○	×	○
教師がどのようにして誤りを直すべきか	明示的フィードバック	誤りがあったことの指摘	○	誤りの部分の強調・正解文の繰り返し		×	○	×	○
		誤りのタイプの指摘	○			○	○	×	○
		誤りの場所の指摘	○			○	○	×	○
		誤り文と正解文を提示し、比較させる	○			○	○	×	○
		正確文の提示	○			○	○	学習者が望む時のみ ○	○
		説明	○			○	○	×	○
	暗示的フィードバック		×			×	×	×	×
	明確化要求		○			×	×	×	×
	非理解を示すフィードバック		○			×	×	×	×

IS YOUR INVESTMENT ADVICE A LOAD OF BULL?
BANNER

More than ever, in turbulent markets, you need sound, independent advice to ensure your money is working hard for you.

BY KEEPING a cool head, thinking beyond any short-term crisis and adopting a disciplined approach to your investments, you can take advantage of volatile markets and ultimately profit from market fluctuations.

Put simply, in a volatile market many good stocks become bargains. That means you buy more for less.

Banner's Offshore Non Contractual Investment Plan does just that – building wealth even in times of volatility through what's called dollar cost averaging. It's a prudent investment strategy whereby your regular contributions are put to work across a range of diverse markets with a medium to long-term goal of maximizing returns.

Because it's non-contractual, you get a truly flexible investment plan.

That's the beauty of it. No contract means guaranteed investment freedom with penalty-free access to your cash at any time.

Banner offers a range of products carefully tailored to the needs and concerns of our clients. For those investors who would prefer to invest larger lump sums, our recently launched Paradigm Fund is a proven performer. The strategy underpinning this fund has made money every month since inception in May 1993.

Whatever investment plan you choose, you benefit from Banner's expertise built up through 20 years of experience in Japan.

And your investments are monitored and analysed in real time, utilising the extensive resources of *Omega Research Analytics, Investment Data Services* and *Bloomberg*.

To find out more call Banner on 03 5724 5100.
We won't baffle you with bull.

Banner Overseas Financial Services Japan KK
03 5724 5100 Fax 03 5724 5300
Email banner@gol.com
Website http://www.bannerjapan.com

Articles

Using a Commercially Produced Proficiency Test in a One-Year Core EFL Curriculum in Japan for Placement Purposes

Brent Culligan
Seigakuin University

Greta Gorsuch
Mejiro University

EFL program administrators have two general testing options for placement of students: commercially produced proficiency tests or locally developed tests. This study focuses on the use of a commercially produced proficiency test (the Secondary Level English Proficiency® test) for student placement in a core EFL program at a private junior college and university in Tokyo. The research was conducted to judge the degree to which the use of the SLEP® test was appropriate for student placement purposes. Pre- and post-test results for 538 students were analyzed for item facility, item discrimination, and item difference indices. It was found that the test did not appear to "fit" the students nor the program. The authors urge the adoption of supplemental placement procedures as well as the development of more program-sensitive tests.

学習者の能力に応じた英語のクラス分けを行う際には、一般に普及している英語能力テストを実施してその得点をもとに行う場合と、各英語コースが独自に開発したテストを実施してその得点をもとに行う場合とが考えられる。本論では、日本の私立短期大学・大学において Second Level English Proficiency Test (SLEP)を用いて英語のクラス分けを行ったケースについて論じる。

本研究の目的は、プレースメント目的でのSLEPの有用性を検証することである。538名に事前・事後テストを実施し、項目分析を行った結果、SLEP®本研究被験者・英語コースのいずれにも適合しないことが判明した。この結果をもとに、より妥当なプレースメントを行うためにテストに併せて実施すべき補足的な事柄と、当該英語コースにより適合したプレースメント・テストの開発について言及する。

EFL program administrators have two general testing options for placement of students: commercially produced proficiency tests or locally developed tests. However, surprisingly little research has been published on the use of commercially produced proficiency tests for student placement in such programs and only a few researchers have published accounts of local placement test development in ESL programs for which the test has been written, piloted, and/or revised by on-site developers (Brown, 1989; Wall, Clapham & Alderson, 1994). This study will describe the use of one commercial test, the Secondary Level English Proficiency® for student placement in a core EFL program at a private junior college and university in Tokyo. The main focus of the research is to assess the degree to which the use of the SLEP® test is appropriate for placement purposes in the program. We seek to determine how appropriately it places students and how well the test "matches" the program goals and objectives. A second interest is to suggest methodology for other researchers to investigate the appropriateness of commercially produced proficiency tests used for student placement in their programs.

"Locally" Developed Placement Tests

"Local" placement tests, if developed along the lines of sound testing principles, have two important advantages. First, such placement tests can be piloted, analyzed, and then revised freely—the type and length of the test need only be limited by the skills of the local test development team and the teachers in the program. Second, such a test can be linked with the curriculum. This second advantage is strongly desirable. In Brown's words, "a placement test must be . . . specifically related to a given program, particularly in terms of the relatively narrow range of abilities assessed and the content of the curriculum" (1996, p. 12). This aspect of test validity is known as content validity. It is the notion that the test content should reflect the content of the curriculum or course it is being used in (Alderson, Clapham, & Wall, 1995; Bachman, 1990; Brown, 1990; Brown, 1995; Brown, 1996; Oller, 1979).

However, these advantages only hold if tests are developed using sound testing principles, including creating test item specifications and item banks, piloting the test, analyzing the test items and the statistical parameters of the test, and then revising the test to improve it on a continuous basis (Alderson et al., 1995; Brown, 1996; Davies, 1990; Henning, 1987). The local test developers would also have to estimate the reliability of the test, determining whether the test was measuring students' traits consistently (Alderson et al., 1995; Brown, 1996; Heywood, 1989; Hughes, 1989; Weir, 1993). Finally, the test developers would have to develop various arguments for the validity of the test. For example, placement decisions could be correlated with students' later achievement in their classes or with the appropriateness of the students' initial placement (Hughes, 1989; Wall et al., 1994).

Developing any sort of test is an arduous process requiring time and adequate knowledge of testing principles. Weir (1993, p. 19) notes that local test development requires group effort. However, having a group of informed and committed test developers in a program is sometimes not possible and administrators and/or teachers in ESL/EFL programs often elect to purchase commercially produced proficiency tests for placement purposes.

Commercially Produced Proficiency Tests

Using commercially produced proficiency tests in a language program has several advantages, the foremost being convenience. As many local test developers will attest, it may take months of committed, enlightened effort to produce a minimally reliable test (Griffee, 1995). Another advantage is economy. For a reasonable sum, programs can purchase testing packages such as the SLEP®. Such packages also include evidence supporting the reliability of the test (Gorsuch, 1995), since testing companies have the resources to make generally reliable tests and to offer well-organized information regarding the valid use of their tests.

An additional reason is ease of administration and scoring. In very large programs such as the one discussed in this study (748 students), it may be impossible to administer tests in which students are interviewed and rated or in which students' writing samples are rated. In such large programs, the number of students may necessitate the use of a paper-and-pencil test, which is the form taken by commercially produced proficiency tests. Finally, such tests may have high face validity in the eyes of students and administrators; commercially produced tests are characterized by professionally laid out and printed pages and high quality tape recordings. The SLEP® test offers an additional advantage. The makers of the test, ETS®, have developed a chart that test administrators can use to estimate students' TOEFL® scores based on their SLEP® scores. That can be valuable in programs in which administrators and/or teachers are anxious to "prove" the value of the program to other interested parties.

However, the literature regarding the use of various kinds of tests for student placement indicates that proficiency tests are a second choice, and even then only in specific kinds of situations. For example, Bachman (1990) suggests the use of proficiency tests for placement when:

1. the students to be tested vary widely in terms of background and language ability;
2. the learning objectives of a program are not clearly specified; and
3. levels of students are known to vary widely from year to year, making the use of a locally developed test normed on one sample of students problematic.

Brown partially agrees: "If a particular program is designed with levels that include beginners as well as very advanced learners, a general proficiency test *might* (italics in the original) adequately serve as a placement instrument." Brown also cautions, "However, such a wide range of abilities is not common . . . in programs" (1996, p. 13).

Yet in most tertiary level EFL programs in Japan the students' second language learning experiences and abilities do not vary widely. Students in these programs have had six years of formal EFL education using similar textbooks and instructional practices. Furthermore, many colleges and universities in Japan are revising their EFL curricula, and have developed program-specific learning goals and objectives. Is the use of commercially produced proficiency tests for placement purposes appropriate for such schools?

As noted, administrators in ESL/EFL programs often choose to use commercially produced proficiency tests for student placement, yet this decision may be problematic. In Brown's words, "Each [placement] test must be examined in terms of how well it fits the abilities of the students and how well it matches what is actually taught in the classrooms" (1996, p. 13). Otherwise students may be placed in class levels based on a test that makes no comment on the curriculum in which the students are enrolled (Brown, 1990). The potential for inappropriate placement can become all too real in such a situation. (For additional cautions concerning the use of proficiency tests for placement, see Brown, 1995; Henning, 1987; and Hughes, 1989.)

Program administrators thus have the difficult choice of using a commercially produced proficiency test which may not be appropriate for placement of their students or they can expend a massive amount of effort writing their own tests. In the end, however, locally written tests may be no more appropriate or reliable than a commercially produced proficiency test. Another option may be to use a commercially produced proficiency test as a stepping stone towards developing a locally written placement test, as will be described below.

Research Focus

This study estimates the extent to which the SLEP® proficiency test is suitable as a placement test for a core English program at a Japanese university. We will address three questions. First, how well does the SLEP® test "fit" the students in the program? Second, how well does the SLEP® test "fit" the goals and objectives of the program? And third, what steps can be taken to improve placement decisions in the program? In answering these questions, we will outline the minimal steps that should be taken to determine the validity of such tests for student placement in tertiary level EFL programs, if reliable and valid "local" tests cannot be developed.

Research Questions

1. What items on the SLEP® test discriminate effectively between high and low scoring students?
2. Will selective scoring of the SLEP® test produce more effective placement of students?
3. To what extent will items from the first and second test administration with high difference index values match the stated goals, objectives, and syllabus of the program[1]

Method

Subjects

The majority of the 748 first-year students enrolled in the university and junior college divisions of the English program during the year of the study were recent graduates from Japanese high schools and were approximately 18 years of age. The students were predominantly of Japanese nationality, with the exception of three Korean students and one Chinese student in the university division. There were 310 males and 87 females in the university division of the program, while the 380 students in the junior college division were all female. In addition, there were seven second-year students in the program who were repeating their first-year English requirements.

The university students were drawn from three majors: Political Science and Economics (268 first-year students), American and European Culture (65), and Early Childhood Education (64). Students in the junior college division majored either in English Literature (180 first-year students and three second-year students) or Japanese Literature (200 first-year and four second-year students).

Material

Two sets of materials were used in this study: the SLEP® test and the core English program goals, objectives, and syllabuses (see Appendix).

SLEP®

The SLEP® test was developed by the Educational Testing Service (ETS®) in 1980, using over 6,000 non-native English speaking secondary school students in the US and in "foreign countries" as its norming population (ETS®, 1991, p. 8). In the words of ETS®, it is a proficiency test and "a measure of ability in two primary areas: understanding spoken English and understanding written English" (ETS®, 1991, p. 7). Further, it is "helpful in evaluating ESL teaching programs and making placement decisions" (ETS®, 1991, p. 7). It is not an aptitude or achievement test.

The SLEP® test currently has three equivalent forms. Students taking the test have a test book and an answer sheet for marking answers. The reported reliability coefficient of the SLEP® is .94 for the listening subtest, .93 for the reading subtest, and .96 for the entire test (ETS®, 1991, p. 9). The SLEP® test is designed to be locally scored, either using a two-ply pressure-sensitive answer form, or an optical recognition form. Scoring here was done using the optical recognition forms and a scoring machine.

The test is made up of a listening section and a reading section, each with 75 multiple choice items. The listening section has four subsections, made up of four different types of multiple choice items. In Form 1, the first listening subsection ("1Pic") asks the students to look at a photograph in the test book and then listen to four sentences on a tape. On their answer sheet the students mark the sentence best describing the photograph. There are 25 items in the "1Pic" subsection. The second listening subsection ("Dict") asks the students to read four sentences in the test book and listen to a sentence recorded on the tape. The students mark the sentence in the test book that is the same as the one on the tape. There are 20 items in the "Dict" subsection.

The third listening subsection ("Map") has 12 items based on an illustration representing a bird's-eye view of a small town. The students identify the buildings and streets on the map and the locations of four cars on the streets. The students then hear short conversations between various adult North Americans on the tape and must surmise in which car the conversation is taking place. The "Map" subsection assumes the cars in the illustration are driven on the right hand side of the road.

The fourth listening subsection ("Conv") has 18 items regarding a North American high school. The students hear several short conversations between adult and teen-age North Americans on the tape. After

Table 1: Summary of Sections and Subsections of SLEP® (Form 1)

Listening Section Subsections	Number of Items	Time Allowed
1Pic	25	
Dict	20	
Map	12	
Conv	18	
		45 minutes

Reading Section Subsections	Number of Items	Time Allowed
Cart	12	
4Pics	15	
Cloze	22	
RP1	18	
RP2	8	
		45 minutes

each conversation, the students hear one or two questions about the conversation and select the correct answer from written items in the test book. The entire listening test with the four subsections takes approximately 45 minutes to complete.

The reading section, which ETS® claims tests grammar and vocabulary, also contains four subsections with four types of multiple choice items. The first reading subsection ("Cart") presents a cartoon illustration in which several people have "thought bubbles" above their heads, each illustrating a different point of view of a particular event. For each item, students read two or three sentences and then match the item to the "thought bubble" of one of the people in the illustration. There are 12 items of this type. The second reading subsection ("4Pics") asks the students to read a sentence, then match it to one of four illustrations which best describe it. There are 15 items of this type.

The third subsection is a short modified cloze reading passage ("Cloze"). For each missing word the students choose one of four possible answers. There are 22 items. The fourth reading subsection ("RP1") contains questions about the preceding passage; the students choose the best answer to the question from four choices. There are 18 items. There are three such modified cloze passages with three sets of questions. Finally, the fifth reading subsection ("RP2") presents a reading passage (without cloze) and eight multiple choice questions about it (eight items).

The students are given 45 minutes to complete the reading test.

See Table 1 for a summary of the tests and subsections of Form 1 of the SLEP® test.

Program Curriculum

In early 1993 two special committees at the university were formed to revise the EFL curriculum. The goal was the creation of a multi-level core EFL program for all first-year university and junior college students, to be implemented at the start of the 1996 academic year. The curriculum design process included administration of a Japanese-language needs analysis questionnaire to 2,067 lower and upper class students at the school in early 1995, numerous in-service lectures conducted by faculty and non-faculty expert/informants over a three year period, readings from the *ACTFL Proficiency Guidelines* (Buck, 1989), and individual study and reflection on the part of the committees' members.

During the period of this study, the program had three levels: A level (high), B level (intermediate) and C level (remedial), corresponding to intermediate/high, intermediate/mid, and intermediate/low levels on the speaking portion of the *ACTFL Proficiency Guidelines* (Buck, 1989). First-year students in the university division attended two 90-minute classes per week for 26 weeks in the core English program, amounting to 78 hours of instruction in one academic year. English Literature majors in the junior college division also received 78 hours of instruction in one academic year, while Japanese Literature majors received 39 hours of instruction given only in the first semester.

Within each level, general goals concerning English proficiency and vocabulary were set, as were objectives describing more precise learn-

Table 2: Recommended Textbooks

Level A
    Atlas II (Nunan, 1996)

Level B
    Atlas I (Nunan, 1996)
    Interchange I (Richards, Hull & Proctor, 1990)
    New Person to Person Book 2 (Richards, Bycina & Kisslinger, 1996b)

Level C
    New Person to Person Book 1 (Richards et al., 1996a)
    First Impact (Ellis, Helgesen, Browne, Gorsuch & Schwab, 1996)

ing outcomes (see Appendix). These goals and objectives resulted in a series of notional/functional syllabuses stressing a communicative approach to language learning. Although objectives for developing students' communicative reading and writing skills were articulated, the program was mainly designed to promote oral/aural skills development.

Based on the program objectives, a selection of textbooks was made for teachers to choose from for use in their classes. (See Table 2.)

In line with goals concerning vocabulary development, a number of learning objectives were specified (see Appendix). After considering materials such as the *Longman Language Activator* (1994), *A General Service List of English Words* (West, 1953) and *A University Word List* (Nation, 1990), a "master vocabulary list" of 3,000 words was compiled using the *Cambridge English Lexicon* (Hindmarsh, 1990), the *Longman Dictionary of Contemporary English* (1995), and the *Cambridge International Dictionary of English* (1995). Vocabulary was broadly sequenced according to frequency to correspond to Levels A, B, and C.

Twenty-five words per week were integrated into the syllabus. Program teachers created weekly vocabulary worksheets based on the 25 words, including crossword puzzles, definition matching, and cloze exercises. The teachers collected the worksheets periodically for correction and comment as formative assessment. Lead teachers assigned to the levels wrote vocabulary quizzes which were given every three weeks to test the students' progress. The vocabulary quizzes contained 25 items taken from the 75 words the students had been studying for the previous three weeks.

Procedure

At the beginning of the 1996 academic year 748 junior college and university students in the program took the SLEP® test Form 1, both listening and reading, for placement purposes. This administration will be referred to as the "pre-test." Nine months later, in January, 1997, 487 students were administered the same Form 1 test for purposes of program evaluation. This is termed the "post-test." The 210 students in the Japanese Literature program did not take the post-test at the same time as the other students because of different degree requirements. Therefore, their scores were not included in this study, nor were those of the 51 university students who failed to take the post-test. Thus, pre-test and post-test scores of only 487 students were used in the analysis.

Data Analyses

To determine which test items discriminated effectively between high and low scoring students (the first research question), the pre-test scores for 487 students on all items of the SLEP® test were entered into a spreadsheet program and were subjected to an item discrimination analysis (ID), a norm-referenced item statistic. According to Brown (1996, p. 66), ID analysis of test items "indicates the degree to which an item separates the students who performed well from those who performed poorly." The ID was calculated for each test item by subtracting the item facility (IF_{lower}) of the students scoring in the lowest third of the test overall from the item facility (IF_{upper}) of the students scoring in the highest third of the test overall. Item facility (IF) is the proportion of students who answered a particular item correctly. For example, if six out of ten students correctly answered an item, the IF would be .60.

Generally speaking, test administrators expect students who score highly on the test overall to also score highly on individual test items. Conversely, administrators expect students with low scores on the test overall to score poorly on most of the individual items. However, the opposite may happen; students who score highly overall may do poorly on individual items. Such items may be poorly constructed, ambiguously worded, or simply too difficult for the students. It is those items that are thought not to discriminate effectively between high and low scoring students and are thus likely to have low item discrimination (ID) values. According to Ebel (as cited in Brown, 1996, p. 70), test items with ID values of .40 and above are considered "very good" items, those with ID values of .30 to .39 are thought to be "reasonably good," and those with ID values of .20 to .29 are "marginal" items, usually "needing improvement." For this study, we looked for items with ID values of .20 and over.

To address the second research question, the high ID items were identified and were taken out of the rest of the data, creating a "high ID" data set. Thus two data sets were analyzed, the original data set with all the items included, and the "high ID" data set,in order to calculate the means, standard deviations, reliability estimates, and standard errors of measurement. This was done to see which data set yielded the more reliable information for placing students appropriately.

To answer the third research question, pre-test scores on individual test items for 487 students were compared to their matching post-test scores using a criterion-referenced test statistic, the difference index (DI) (Brown, 1996, p. 80). DI was calculated by subtracting pre-test item facility (IF) for each item from post-test IF for each matching item. Thus, if students did better on particular items on the post-test, the DI for those items had a positive value. Items with DI values of .10 or over were examined in light of the stated goals, objectives, and syllabuses of the program. In particular, we looked for any patterns in students' improvement in terms of SLEP® tests (listening and reading) and subtests ("1Pic," "Dict," "Map," etc.). We wanted to see the extent to which the SLEP® test "matched" the program goals, objectives, and syllabus statements.

We would like to note here that although we used the goals, objectives, and syllabuses of the program to gauge the degree of fit between the program curriculum and the SLEP®, the implementation of the goals and objectives was not investigated. This issue is central to the whole question of defining what a curriculum is and what it does (i.e., program evaluation) (Holliday, 1992; Snyder, Bolin & Zumwalt, 1992; White, 1988). Our study, we feel, constitutes only one part of such a program evaluation. However, in Brown's (1995) model of curriculum development the establishment of objectives is followed by testing, and is then subject to evaluation. This first step is the limited scope of our study.

Results

Upon analysis of the pre-test data, we found that less than half of the items had an ID of .20 or higher, the minimum level thought acceptable for effective discrimination (Ebel cited in Brown, 1996). See Table 3 below.

Table 3: Pretest Items with ID of .20 and Above

Section	Subsection	Items with ID of .20 and Above	Total Items in Subsection
Listening	1Pic	16	25
Listening	Dict	20	20
Listening	Map	5	12
Listening	Conv	1	18
Reading	Cart	10	12
Reading	4Pics	6	15
Reading	cloze	4	22
Reading	RP1	2	18
Reading	RP2	2	8
Totals		66	150

The first research question asked which items on the SLEP® test discriminated effectively between high and low scoring students. Of the 66 items with "acceptable" IDs, 42 were listening section items and 24 were reading section items. The test thus appears to have discriminated better for listening than for reading. The remaining 84 items had an ID of .19 or below and, by Ebel's standards (as cited in Brown, 1996), were not useful for discriminating between high and low scoring students.

In answering the second research question, two data sets were created to see whether selective scoring of the SLEP® test would result in more effective placement of students. The "original data set" included data for all 150 items in the SLEP® test, whereas the "high ID data set" included data for only those 66 items that were found to have an ID of .20 or over (see Table 3 above). Comparisons of descriptive statistics on the two data sets are given in Table 4. Also included are KR-20 internal consistency estimates for the two data sets.

Table 4: Comparisons of Original Data Set and High ID Data Set

	Original Data Set	High ID Data Set
K	150	66
M	69.36	39.60
SD	12.38	9.05
high	107	61
low	32	11
range	76	51
KR-20	0.81	0.84
SEM	5.46	3.62

The standard error of measure (SEM) of the high ID data set is substantially lower than that of the original data set, whereas the KR-20 internal consistency estimate is somewhat higher for the high ID data set. These results indicate that selective scoring of the SLEP® test would most likely result in more effective placement of students in the program.[2]

Finally, to answer the third research question, regarding whether items from the first and second test administration with high difference index values match the goals and objectives of the program, pre-test and post-test data were compared to calculate the difference index (DI) for each item, thus estimating students' gain scores on particular items. Items with a DI of .10 or better by SLEP® test subsection are shown in Table 5.

Thirty-one of the "high DI" items were in the listening section and 16 were in the reading section. Four subsections had six or more items with high DIs, four subsections had items with low DIs, and one subsection had items with DIs of zero. Each of the subsections will be analyzed below and compared to the goals, objectives, and syllabuses of the core English program in order to understand the extent to which the items in the subsections "fit" the curriculum.

Table 5: Items with DI of .10 and Above

Section	Subsection	Number of High DI Items	Total Items in Subsection
Listening	1Pic	13	25
Listening	Dict	15	20
Listening	Map	2	12
Listening	Conv	1	18
Section Total		31	75
Reading	Cart	2	12
Reading	4Pics	0	15
Reading	Cloze	6	22
Reading	RP1	6	18
Reading	RP2	2	8
Section Total		16	75
Total		47	150

As shown in Table 5, students showed gain scores on 13 out of 25 items in the "1Pic" subsection, which focuses primarily on meaning; students see a picture, hear four statements, and then decide which statement matches the picture. While the goals and objectives for the core English curriculum cannot be explicitly matched with the subsection in terms of content, the goals and objectives statements for Programs A, B, and C (see Appendix) calls for students to learn how to "ask and answer questions" in a variety of settings. The goals and objectives statement for Program A mentions that students should learn to "understand and respond to extended discourse." If teachers created classroom activities based on these goals and objectives, perhaps these activities gave the students meaning-focused listening practice, either through pair work, completing listening activities in textbooks, or listening to extended lectures in English.

On the "Dict" listening subsection of the test, students showed high gain scores on 15 out of 20 items (see Table 5). Items in this subsection were more oriented to form than meaning. Students had to listen to a statement and match it with one of four written statements in the textbook. The connection between items of this type and the core curriculum is more tenuous and indirect. Only the Program A goals and objectives statements concerning the improvement of students' note-taking ability can be directly related to this subsection. Note-taking practice requires accuracy in listening. In addition, all the textbooks listed in Table 2 utilize tape-recorded listening activities which focus on accuracy in listening. We speculate that activities designed to meet the meaning-focused goals and objectives for listening had a "spill over" effect which improved students' accuracy in hearing and identifying English forms. Another possibility is that activities designed to fulfill the goals and objectives related to improving students' reading helped students to improve their scores in this listening subsection. Such test items require more reading skill than would at first seem apparent. In order to answer the items, students must "race ahead" of the tape and read the four answer statements quickly and accurately before the test statement is played on the tape. After the statement is played, the students must quickly read the answers again to evaluate which one is being said. It may be that students' reading practice in the core English program helped them read the answer choices on this subsection of the test more efficiently.

On the "cloze" reading items in the test (see Table 5), students showed gain scores on only 6 out of 22 items. While some of the cloze items tested vocabulary, many of them seemed to test the students' judgments of correct word morphology. Students were given four versions of the same verb or adjective and had to choose the most appropriate one. Of these six items, two indicated an increase in vocabulary knowledge, two showed gains in students' morphological discrimination, and two showed an increase in students' ability to choose correct function words, such as referents. The students' relative improvement on the six items may be partly due to the program's weekly vocabulary worksheets mentioned above. The vocabulary worksheets took a variety of forms, including cloze exercises and definition matching games, but presented the vocabulary items in the morphological form required for the correct answer. We speculate that students received input that promoted an inductive understanding of correct word morphology and syntactic structure on the relevant items in the SLEP® test.

The students showed an improvement on 6 out of 18 items (see Table 5) on the "RP1" subsection, and this seemed to have an indirect relationship to the goals and objectives of the program. The items in this subsection required the students to infer meaning. It is possible that through meaning-focused listening and reading activities, designed and used in accordance with the goals and objectives of the program (i.e., "understanding extended discourse," "reading written materials for information," "carrying on simple face to face conversations"), the students' ability to answer meaning-focused test questions improved.

As shown in Table 5, students showed little or no gain on five subsections: "Map," "Conv," "Cart," "4Pics," and "RP2." There are several explanations for this. Students already had fairly high scores on the "Cart" and

"4Pics" subsections on the pre-test. Thus, there was not much room for improvement. The "Cart" subsection pre-test item facilities (IFs) for 10 out of 12 items were .60 or over. In the "4Pics" section, 10 out of 15 items had pre-test IFs of .60 or over. These high values suggest that the items in the two subsections were generally easy for the students.

The small gains shown by students in the "Map" and "Conv" subsections probably have different causes. The students' pre-test IFs for most of the items in these subsections were low and remained so in the post-test. We feel that the two subsections were simply too difficult for these students because they were culturally inappropriate. Both the "Map" and "Conv" subsections assumed experiences that first-year Japanese college students are unlikely to have had. For example, the "Map" subsection assumed that the testees had done extensive car travel, or could drive, particularly on the right side of the road. However, most young Japanese do not get driver's licenses until they are 20 years old and then drive on the left hand side of the road.

Similarly, the "Conv" section assumes students are familiar with the duties of administrative personnel in American high schools. However, there is no guarantee that administrative counterparts in Japan handled the same duties, or even that there are such administrators in Japanese high schools. We feel that regardless of the language learning support students received in the program, the "Map" and "Conv" subsections presented unfamiliar concepts. Thus, students could not effectively demonstrate their learning through these two subsections.

The modest gains shown on the final subsection, "RP2" may have been due to students' unfamiliarity with the genre of fictional short reading. Many students are familiar with expository written English since this makes up the bulk of the reading presented in high school textbooks. However, they may be less familiar with stylistic devices and imagery used in fiction. The goals and objectives statements for program levels A, B, and C (see the Appendix) allude to reading in functional terms. In level A for example, students are asked to read easy "academic" materials. Students in levels B and C are asked to read "public transport schedules," "newspaper articles," and "notes from the teacher." The program is not intended to promote students' reading of literary works in English. Thus, this particular subsection is not really connected to the program, either in content or in terms of what activities students are asked to do.

Discussion

According to Bachman (1990, p. 238), test validity is not an abstract notion. Rather, test validity must be considered in the context of the inferences that teachers or program administrators plan to make from the students' test results. Thus, in a situation where a commercially produced proficiency test is used to place students in different levels in a program, we need to answer the question of whether the test is valid for this purpose, i.e., whether the test "fits" the students and "fits" the program.

There are a number of reasons why the SLEP® test does not appear to be valid when used for placement of students in the core EFL program described in this study. First, we found that only 66 out of a total of 150 items on the test discriminated between high and low scoring students. The result was a standard error of measure of 5.46 (see Table 4), indicating a good deal of "looseness" around the cutoff points used to decide whether students should be placed in the A, B, or C levels of the program.

Second, the SLEP® test does not estimate oral ability, although an aim of the program is to increase students' oral skills. This alone constitutes a mismatch between the test and the program. We were able to make only indirect comparisons between the program's listening and reading goals and objectives and various SLEP® subsections, but these comparisons were at best speculative. The SLEP® test, therefore, does not seem to "fit" this particular program.

However, as discussed, administrators and/or teachers often elect to use commercially produced proficiency tests for placement in a program with defined goals and objectives. In our particular situation, the large number of students (748) made oral testing for placement purposes prohibitively difficult. Also, as this was the first year the core EFL program was in place, there was no possibility of developing a local paper-and-pencil test more suited to the students and to the program. We strongly hope that as the program continues the administrators and teachers will consider developing a reliable and valid local test or will develop placement procedures to supplement the SLEP® test. The data that we have gathered through this study can be of some assistance. For example, item types from the SLEP® test that consistently produce high gains and/or high discrimination can be used as models for item writing for the local placement test.

We suggest that the SLEP® test, if scored with all 150 items, is problematic for placement of the students in the program described above. We therefore recommend that the test be scored selectively, using only the 66 high ID items. By selectively scoring the SLEP® test, the program administrators may be able to obtain more effective placement of students by reducing error variance. Although the number of test items counted toward the total score would be reduced, the reliability of that score would increase. By scoring only the 66 items with high IDs, the SEM dropped from 5.46 to 3.62. The SEM is best conceived as "a band around a student's score within which that student's score would probably fall if the test were administered to him or her repeatedly" (Brown, 1996, p. 206). We interpret this to mean that on the total test, the true score of a student who got a raw score of 70 could actually range from plus one SEM to minus one SEM 68% of the time, from 65 to 75. For the remaining 32%, the measurement error could be greater. This can result in the misplacement of "borderline" students. Reducing the SEM by selectively scoring the pre-test would reduce misplacement.

Continual assessment of the test items, such as we did in this study, will provide much needed "tuning" for educational institutions using proficiency tests, whether locally developed or commercially produced. With this in mind, we must assert that the results of this study cannot be used as justification for using portions of the SLEP® test in any other Japanese institutional setting. Only with continual monitoring of the results on an item-by-item basis can valid inferences be made using the SLEP®, or any other test, for a particular setting. As testing situations change, so must the assessment of the validity of the tests used.

Acknowledgments

The authors would like to thank J.D. Brown for his instruction and encouragement, and Dale T. Griffee, William Kroehler, and the three anonymous JALT Journal *reviewers for their insightful comments. Thanks are also due to the administrators, teachers, and students in the core English program at Seigakuin University.*

Brent Culligan is a full time instructor at Seigakuin University. A doctoral candidate at Temple University, Tokyo, he is interested in second language vocabulary acquisition and testing.

Greta Gorsuch teaches full time at Mejiro University, and is former editor of *The Language Teacher*. She is a doctoral candidate at Temple University, Tokyo, and is interested in testing, instruction processes, and education policy evaluation.

Notes

1. One of the reviewers objected to our use of this research question. She/he felt quite rightly that a multiple choice listening and reading test (such as the SLEP®) could not be considered appropriate for use in a program designed to promote students' oral/aural skills. However, we felt we needed to retain this research question. As stated earlier, one of our purposes is to suggest a method for readers to judge commercially-produced proficiency tests used for placement in their own programs. We feel that research question three presents a useful tool for relating the test to the program.
2. One reviewer suggested that in order to confirm our claim we would have to assess the students' progress over a semester to gauge the appropriateness of their placement using the high ID data set. While we feel this is a

cogent point, we also feel that in practical terms this would be difficult to carry out. Such an assessment would require comparing a control group (students placed using the original data set) to an experimental group (students placed using the high ID data set). Even if this or a time series study had been done, we would have to consider that these students' progress could be due to a multitude of factors and could not necessarily be attributed to appropriateness of student placement.

References

Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Brown, J.D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 23 (1), 65-83.

Brown, J.D. (1990). Where do tests fit into language programs? *JALT Journal*, 12 (1), 121-140.

Brown, J.D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle & Heinle.

Brown, J.D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Buck, K. (Ed.). (1989). *The ACTFL oral proficiency interview: Tester training manual*. New York: The American Council on the Teaching of Foreign Languages.

Cambridge international dictionary of English. (1995). Cambridge: Cambridge University Press.

Davies, A. (1990). *Principles of language testing*. Oxford: Blackwell.

Educational Testing Service®. (1991). *SLEP® test manual*. Princeton, NJ: Author.

Ellis, R., Helgesen, M., Browne, C., Gorsuch, G. & Schwab, J. (1996). *First impact*. Hong Kong: Longman Asia ELT.

Gorsuch, G.J. (1995). Tests, testing companies, educators, and students. *The Language Teacher*, 19 (10), 37, 39, 41.

Griffee, D.T. (1995). Criterion-referenced test construction and evaluation. In J.D. Brown & S.O. Yamashita (Eds.). *JALT applied materials: Language testing in Japan* (pp. 20-28). Tokyo: Japan Association for Language Teaching.

Henning, G. (1987). *A guide to language testing*. Boston: Heinle & Heinle.

Heywood, J. (1989). *Assessment in higher education* (2nd ed.). Chichester, MA: John Wiley and Sons.

Hindmarsh, R. (1990). *Cambridge English lexicon*. Cambridge: Cambridge University Press.

Holliday, A. (1992). Tissue rejection and informal orders in ELT projects: Collecting the right information. *Applied Linguistics*, 13, 403-424.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Longman dictionary of contemporary English (3rd. ed.). (1995). London: Longman.

Longman language activator. (1994). Harlow, Essex: Longman House.

Nation, I.S.P. (1990). A university word list. In I.S.P. Nation (Ed.). *Teaching and learning vocabulary* (pp. 235-239). New York: Newbury House.

Nunan, D. (1996). *Atlas I.* Boston: Heinle & Heinle.

Nunan, D. (1996). *Atlas II.* Boston: Heinle & Heinle.

Oller, J.W., Jr. (1979). *Language tests at school.* London: Longman.

Richards, J., Bycina, D. & Kisslinger, E. (1996a). *New person to person, book 1.* Oxford: Oxford University Press.

Richards, J., Bycina, D. & Kisslinger, E. (1996b). *New person to person, book 2.* Oxford: Oxford University Press.

Richards, J., Hull, J. & Proctor, S. (1990) *Interchange I.* Cambridge: Cambridge University Press.

Snyder, J., Bolin, F. & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.). *Handbook of curriculum research* (pp. 402-435). New York: MacMillan.

Wall, D., Clapham, C. & Alderson, J.C. (1994). Evaluating a placement test. *Language Testing,* 11(3), 321-344.

Weir, C. (1993). *Understanding and developing language tests.* New York: Prentice Hall.

West, M. (1953). *A general service list of English words.* London: Longman, Green and Co.

White, R.V. (1988). *The ELT curriculum: Design, innovation and management.* Oxford: Blackwell.

(Received January 11, 1998; revised June 13, 1998)

Appendix

Goals and Objectives for Levels A, B, and C

Goals and Objectives for Program A (intermediate-high)

Course Overview: The purpose of this course is to prepare students to understand and to respond to extended discourse such as lectures, TV and radio talks, to make simple presentations, and to narrate in the past.

Goals	Objectives
1. Increase mastery of vocabulary and idioms in order to expand the range of situations in which students can function in English, and in order to gain competency in academic pursuits.	Be able to score at least 80% on a vocabulary test on approximately 3500+ words including the *University Vocabulary* and other high frequency vocabulary items. Be able to score at least 80% on a test of 700 high frequency idioms (including the 500 in Program B).
2. Understand extended discourse.	Listen to and understand simple lectures and speeches in general and academic settings.
3. Ask questions regarding extended discourse; narrate in the past.	Be able to ask pertinent questions regarding lectures and speeches; be able to make presentations such as a report in a seminar; be able to narrate events and experiences in the past.
4. Read written materials of increasing difficulty for gathering information for personal and academic purposes.	Be able to understand simple academic writing and an increasing number of newspaper and magazine articles.
5. Note-taking and academic writing.	Take notes on lectures, write simple reports based on reading materials, taking into consideration citation and bibliographical protocols.

Goals and Objectives for Program B (intermediate-mid)

Course Overview: The purpose of this course is to prepare students to participate in simple conversations about their personal history, leisure time activities, etc., to recognize different registers (politeness, etc.), to listen to simple announcements and use the telephone, to read descriptions of persons, places and events, and to write simple letters or compositions on assigned themes.

Note: Goals and Objectives for Program C are assumed, and if necessary some review of goals and objectives for Program C will be included in Program B.

Goals	Objectives
1. Increase mastery of essential vocabulary and idioms to increase overall mastery of English, and in order to be able to effectively use an English/English dictionary designed for ESL learners.	Be able to score at least 80% on a vocabulary test on 2,500+ word level expanded from the vocabulary list in Program C from such lists as the Key Concepts in the *Longman's Activator Dictionary*; be able to score at least 80% on 500 high frequency idioms (including the 300 in Program C).
2. Be able to ask and answer questions and carry on face-to-face conversations when traveling overseas and in a setting such as a homestay in an English-speaking family.	Ask and give information about travel plans; offer, accept and refuse invitations; explain aspects of one's culture; describe health problems, etc.
3. Be able to read a widening range of written materials for essential information and for enjoyment.	Be able to understand and read public transport schedules, notices and advertisements, and simple newspaper and magazine articles.
4. Be able to convey increasingly complex ideas and information through written English.	Write letters and expanded compositions about daily activities and social activities; write more detailed book reports.

Goals and Objectives for Program C (intermediate-low)

Course Overview: The purpose of this course is to prepare students to be able to introduce themselves, ask and answer simple questions and successfully handle a limited number of interactive, task-oriented and social situations, and to convey and gather basic information through writing.

Goals	Objectives
1. Increase mastery of essential vocabulary and idioms in order to increase overall English ability, and in order to be able to begin using an English/English dictionary designed for ESL learners.	Be able to score at least 80% on a vocabulary test on the 2,000+ word level developed in-house from *West's General Service List, Longman Defining* vocabulary; be able to score at least 80% on 300 high frequency idioms.
2. Be able to ask and answer questions, and carry on simple face-to-face conversations such as self-introductions, ordering a meal, asking directions, making purchases.	Participate in role plays, greet and carry on minimal conversations with native speakers on campus, understand and respond to classroom instructions in appropriate ways.
3. Be able to gather basic information from simple written English instructions.	Become familiar with written English instructions in order to take tests without resorting to the use of Japanese. Be able to read class notices and notes from the teacher. Read simplified graded readers.
4. Be able to convey simple messages through written English.	Write simple answers to questions. Write simple short passages such as self-introductions, everyday activities, plans.

Evaluating Six Measures of EFL Learners' Pragmatic Competence

Ken Enochs
International Christian University

Sonia Yoshitake-Strain
Seigakuin University

This study examines the reliability, validity, and practicality of six measures of cross-cultural pragmatic competence. The multi-test framework used here was developed by Hudson, Detmer, and Brown at the University of Hawaii and consists of six tests which focus on the students' ability to appropriately produce the speech acts of requests, apologies, and refusals in situations involving varying degrees of relative power, social distance, and imposition. These measures have previously been tested on native Japanese learners of English in an ESL context (Hudson et al., 1992, 1995) and on learners of Japanese in a JSL context (Yamashita, 1996). The current study administered these tests to native Japanese learners in an EFL context. Four of the tests proved highly reliable and valid and two of the tests less so. Furthermore, the tests clearly differentiated those students who had a substantial amount of overseas experience from those who had not, a distinction not shown by the students' TOEFL scores.

本論では、6種類のプラグマティック能力測定テストの妥当性・信頼性・実用性を検討する。この6種類のテストは、外国語学習者が習得している、様々な状況下での適切な依頼・謝罪・断わり表現の生成能力を測定するため、ハワイ大学 のHudson, Detmer, and Brownによって開発されたものであるが、これまでにも日本人英語学習者(ESL)や日本語学習者を対象にした研究がなされている。

本研究では、この6種類のテストを日本人英語学習者(EFL)を対象に測定をおこなった。その結果、6種類のうち4種類のテストにおいて高い妥当性と信頼性を確証した。さらに、これらのテストでは、TOEFLでは判別できなかった海外滞在経験の有無を明確に判別することが可能であった。

The notion that language competence involves the ability to produce language that is not only grammatically correct but also appropriate for particular situations has been fundamental to language learning pedagogy and research for decades. According to Mundby (1978), "to communicate effectively, a speaker must know not only how to produce any and all grammatical utterances of a language, but also how to use them appropriately. The speaker must know what to say, with whom, and when and where" (p. 17). A number of linguists over the years (Hymes, 1972; Canale & Swain, 1980; Canale, 1988; Bachman, 1990; etc.) have used the term *communicative competence* to account for the contextual and socio-cultural knowledge that is necessary to use language in real-life situations. Bachman (1990) has suggested that *communicative competence* consists of two interactive components: *organizational competence* to account for grammatical knowledge, and *pragmatic competence* to account for the "capacity for implementing, or executing [organizational] competence in appropriate, contextualized communicative language use" (p. 84).

Deficiencies in pragmatic competence result in what is commonly called pragmatic failure. Thomas (1983) has broadly defined pragmatic failure as occurring "on any occasion the speaker's utterance is perceived by a hearer as different than what the speaker intended should be perceived" (as cited in Hudson, Detmer & Brown, 1992, p. 5). A great deal of research has been directed at defining the causes of pragmatic failure, much of it focused on the inappropriate realization of speech acts. Speech acts are defined as "not an 'act of speech' . . . but a communicative activity . . . defined with reference to the intentions of speakers while speaking and the effects they achieve on listeners" (Crystal, 1991, p. 383).

Three such speech acts that involve very different strategies depending on the culture are requests, refusals, and apologies (Beebe & Takahashi, 1989; Beebe, Takahashi, & Uliss-Weltz, 1990). Furthermore, Hudson et al. (1992, 1995) claim there are different perceptions between speakers of different cultures regarding variables such as *relative power*, *social distance*, and *degree of imposition*. Relative power has to do with the extent to which the speaker's will can be imposed on the hearer. An employer, for example, would have +power over an employee, whereas an employee would have –power with an employer. Social distance refers to the degree of familiarity between the speaker and hearer. For example, speaking with a stranger would involve +distance, whereas speaking with a housemate or co-worker would involve –distance. Finally, the degree of imposition is the right and extent to which the speaker imposes on the hearer. As examples, asking to borrow a dictionary involves –imposition, while asking someone to spend a Saturday helping one to move would involve +imposition.

These three variables, relative power, social distance, and degree of imposition, are considered to be especially significant because "within the research on cross-cultural pragmatics, they are identified as the three independent and culturally sensitive variables that subsume all other variables and play a principal role in speech act behavior" (Hudson et al., 1995, p. 4). Therefore, situations that combine the speech acts of requests, refusals, and apologies with the variables of power, distance, and imposition provide learners with a rich array of pragmatic challenges.

In an effort to determine how pragmatic competence might best be assessed, Hudson et al. (1992) produced six different tests of varying type and method, each involving situations that combine the speech acts of requests, refusals, and apologies with the socio-cultural variables of power, distance, and imposition. They administered these tests to native Japanese students studying English in an ESL context and reported their results in *Developing Prototypic Measures of Cross-Cultural Pragmatics* (1995). Additionally, Yamashita (1996) administered these same tests (translated into Japanese) to a group of second-language learners of Japanese in a JSL context. The current study administered these tests to Japanese students in an EFL context for the purpose of analyzing the results both qualitatively and quantitatively. Yoshitake-Strain concentrated on qualitative analysis and reported her findings in her Ph.D. dissertation, *Interlanguage Competence of Japanese Students of English: A Multi-test Framework Evaluation* (1997), and the present researchers have recently published a preliminary statistical analysis (Enochs & Yoshitake, 1996) on the use of the self-assessment and role play tests in assessing pragmatic competence. The purpose of this investigation is to report on a statistical analysis of the reliability, validity, and practicality of all six tests. The following research questions were addressed:

Research Question 1. How reliable are these test formats for measuring Japanese EFL students' pragmatic competence? Reliability will be determined using internal consistency estimates, measures of inter-rater reliability, and the standard error of measurement (SEM).

Research Question 2. How valid are these test formats? Validity will be determined in terms of content, criterion-related, and construct validity.

Research Question 3. How practical are these test formats?

Method

Participants

The participants in this study were 25 first-year students in the English Language Program (ELP) at International Christian University (ICU) in Tokyo, where both authors were working at the time the data were collected. Most of the students were non-English majors, and all were volunteers who participated in the study during their out-of-class free time. There were seven male and 18 female students, with ages ranging from 18-20, and one 26-year old. The students had started the program in April and were tested in October, having completed the spring term and several weeks of the fall term prior to the test. During both terms, the students' English-language study consisted of approximately nine 70-minute classes per week in a content-based curriculum focused on developing the students' ability in academic English. The students tested were considered to be "average" within the context of the ELP, since they were drawn from the middle of the three placement levels in the program. The TOEFL scores for these students ranged from 423-577 points, with most of the students falling in the 500-539 range. The scores were obtained upon entrance into the university in April.

The overseas experience of the students varied, with many having recently returned from six-week academic English programs at universities in English-speaking countries as part of ICU's Summer English Abroad (SEA) Program. The distribution of the students' overseas experience is broken into three categories (see Table 1). Group 1 had none or very little overseas experience. Those who did have some experience generally gained it through a vacation with their family, which it was reasoned would have had negligible effect on the students' English linguistic and pragmatic competence. The members of Group 2 had spent at least five weeks overseas, generally in homestay situations, and students participating in the SEA Program had been immersed in university summer English-language programs as well. Members of Group 3 had all lived overseas, and were considered to have had a significant amount of exposure to English.

Table 1: Overseas Experience of Subjects

Group	Time overseas	*n*	Comments
1	None or little	8	2 had none, 6 had 2–3 weeks experience, generally in English-speaking countries.
2	5-10 weeks	12	All had experienced some sort of English-language immersion, many through participating in ICU's SEA program.
3	Returnees	5	One to 6.5 years overseas. While only one had lived in an English-speaking country (for 2 years), others had attended international schools in which the language of instruction was mainly English.

Instruments and Administrative Procedure

The six tests administered and evaluated in this study were developed at the Second Language Teaching and Curriculum Center of the University of Hawaii by Hudson, Detmer, and Brown (1992, 1995). These tests were designed as prototypic measures of cross-cultural pragmatic competence. While each of these tests focuses on the three key variables of power, social distance, and degree of imposition in the speech acts of requests, refusals, and apologies, the tests vary in their type and method. The reason for this was to develop "instruments of different types and methods for application across different social variables and speech acts" and reflects the need to determine "the potential differential effectiveness of the instruments" (1995, p. 6). The tests are listed below in the order they were administered in the present study.

1. Self-Assessment Test (SA)
2. Listening Laboratory Production Test (LL)
3. Open Discourse Completion Test (OPDCT)
4. Multiple Choice Discourse Completion Test (MCDCT)
5. Role-play Self-Assessment Test (RPSA)
6. Role-play Test (RP)

For all of these tests, Hudson et al. designed a framework which would evenly distribute various combinations of the attributes they wished to measure. With three different speech acts and eight different combinations of power, distance, and imposition, 24 cells were necessary to represent all combinations of these attributes. These various combinations were randomly reordered and then consistently applied to various task situations throughout the series of tests (see the table in Hudson et al., 1995, p. 10, which shows how these combinations were distributed in their research using tests with 24 different items).

For the RPSA and RP tests, participants performed one series of eight different role play scenarios in which each scenario contained a request, a refusal, and an apology. The socio-cultural variables, however, were similarly distributed in a random fashion. For all of the tests except for the MCDCT, either students or raters indicated on a five-point Likert scale how well they felt the speech act situations had been performed. Details regarding the administration and specific nature of each of these tests follow. For single-item examples of each of the tests, see the Appendix.

Self-assessment test (SA)

The first test administered of the series, this test provided participants with written descriptions of each of the twenty-four speech act situations. After reading each situation, they indicated on a five-point Likert scale how well they felt they could provide an appropriate response in each of the situations. The Appendix shows an example of an apology situation with –imposition, +power, and –distance.

Listening Laboratory Production Test (LL)

This test provided participants with tape-recorded descriptions of the situations to which they provided oral responses. Each description was given twice, and the participants then recorded what they felt was an appropriate response during a one-minute interval following the second listening. Raters then listened to the responses and evaluated each of them using the same five-point Likert scale. The Appendix shows an example of an apology situation with +imposition, -power, and +distance.

Open Discourse Completion Test (OPDCT)

This test was given as a take-home assignment, which participants were given one week to complete. Each participant signed a written pledge that he or she would not receive any assistance on this test. Here, the 24 descriptions of various speech act situations were provided in written form, and the participants were required to provide an appropriate written response to each situation. Raters read the written responses and evaluated each of them using the same five-point Likert scale. The Appendix shows an example of a request situation with +imposition, -power, and +distance.

Multiple-Choice Discourse Completion Test (MCDCT)

This test was also given as a take-home assignment (and participants were reminded of their pledge not to seek assistance). Again, written descriptions were provided of different situations, but this time the participants could choose an appropriate response from among three multiple-choice possibilities, only one of which would be considered fully appropriate by a native speaker of English. Evaluating this test involved giving five points for each correct response (according to a key provided by the test developers), and zero points for either of the incorrect responses. The Appendix shows an example of a refusal situation with -imposition, -power, and -distance.

Role-Play Self-Assessment Test (RPSA)

This test required students to perform the speech act situations as role plays, with a native speaker of English acting as interlocutor. In this test there are just eight different scenarios, but each includes all three speech acts—a request, a refusal, and an apology—with varying degrees of power, distance, and imposition in each situation to mirror the other tests with 24 separate situations. Written descriptions of the role plays (in both English and Japanese) were given to the participants beforehand so they could have a clear understanding of each situation and of what would be expected of them. These role plays were performed in a studio-like room at ICU and recorded on videotape. Immediately after performing each role play, the participants rated on the same five-point Likert scale how well they felt that they had appropriately responded in these speech act situations. The Appendix shows an example used for both the RPSA and RP tests in which all three speech acts were performed in a situation with -imposition, -power, and +distance.

Role-play test (RP)

Using the videotape recordings of the role plays, raters used the same five-point Likert scale to evaluate the appropriateness of each of the 24 speech acts within the eight role plays.

Statistical Analysis

Each of the tests had 24 different items. All of the tests, with the exception of the MCDCT, used 5-point Likert scales, making a total possible score of 120 points. With the MCDCT, 5 points were given for each right answer so a total possible score for this test was also 120 points. These data were initially entered onto a spreadsheet using Excel 5.0. They were then analyzed using Excel and the statistics program SSPS/PC+ Version 4.0.1. Estimates of reliability were conducted through an analysis of internal consistency, inter-rater reliability, and the standard error of measurement. Validity was analyzed in terms of content, criterion-related, and construct validity. The determination of construct validity was made through a principal components analysis, factor analysis, a multivariate analysis and a univariate follow-up statistic of differential groups.

Inter-rater reliability

Three raters were used for each of the tests that required raters—the LL, OPDCT, and the RP test. These were drawn from a pool of raters made up of colleagues and one spouse, a mix of men and women of approximately the same age and educational background. They consisted of five Americans and one Englishman and were all ESL professionals, with the exception of one of the Americans being a journalist. Training involved first an explanation of the speech acts and variables being examined. Raters were then asked to make holistic evaluations of the appropriateness of the students' responses without regard for grammatical accuracy.

Estimates of the inter-rater reliability were first made using the Pearson product-moment correlation coefficients (Pearson r) for different pairings of raters, as can be seen in Table 2.

The highest correlations were clearly between the raters on the RP test, followed by those for the LL test. There was considerably less correlation between the raters on the OPDCT test.

As Brown points out, the number of ratings "can have a dramatic effect on the magnitude of the reliability coefficient" (1996, pp. 203–204). The ratings of the three raters together, then, will tend to be more reliable than a given pair, and "adjusting to find the reliability of larger numbers of ratings taken together would be logical, possible, and advisable" (p. 204). The full tests inter-rater reliability estimates using the Spearman-Brown Prophecy formula[1] can be seen in Table 3. Converted to percentages, the RP test provides an estimated 93% reliability, followed by the LL test at approximately 80%, and the OPDCT test at 49%.

Table 2: Inter-rater Correlation Matrix Using Pearson *r*

	LL test		
	Rater 1	Rater 2	Rater 3
Rater 1	1.0000		
Rater 2	.6428**	1.0000	
Rater 3	.5350*	.5139*	1.0000
	OPDCT		
	Rater 1	Rater 2	Rater 3
Rater 1	1.0000		
Rater 2	.2705	1.0000	
Rater 3	.1590	.3012	1.0000
	RP test		
	Rater 1	Rater 2	Rater 3
Rater 1	1.0000		
Rater 2	.7894**	1.0000	
Rater 3	.8069**	.8413**	1.0000

*$p < .01$

**$p < .001$

Table 3: Inter-rater Reliability Using Spearman-Brown

LL	OPDCT	RP
.7957	.4933	.9296

Results and Discussion

Descriptive Statistics

Table 4 shows descriptive statistics including the mean, standard deviation, minimum, maximum, and range of the scores for 25 students. The TOEFL results reveal a mean of 502 points which is somewhat higher than the Japanese national average of 494. The average mean of the TOEFL subtest scores of 49.48 for Listening, 51.28 for Structure, and 50 for Reading are correspondingly higher but basically parallel to the Japanese national average of 49 for Listening, 50 for Structure, and 49 for Reading (Educational Testing Service, 1995).

As for the six tests designed by Hudson et al. and administered to EFL students in the present study, several of the descriptive statistics are worth noting. Of the two discourse-completion tests, the OPDCT had the highest mean score at 92.48, but the lowest standard deviation at 6.70. This contrasts sharply with the MCDCT which had the lowest mean score at 70, but the second to the highest standard deviation at 14.43. Of the two self-assessment tests, it is interesting to note the relatively high mean score of 86.08 for the SA test, which had the highest standard deviation at 14.59 points. In this test, participants *speculated* on the degree to which they could demonstrate pragmatic competence in particular situations. In comparison, the RPSA had a similarly high standard deviation of 14.31, but a considerably lower mean at 78.88. This score reflects how well participants felt they *realized* pragmatic competence in their role play performances. The substantially lower mean for the RPSA suggests that the participants in this study generally did not feel they had performed as well as they thought they could in these situations.

For the RP test, the mean of the raters' scores was identical to that of the RPSA at 78.88 points, but with a considerably lower standard deviation: 10.53 versus 14.31. There was also a significant variation between the raters of the LL test, ranging from a high of 81.6 to a low of 65.2. Of the individual raters' scores for the three tests which required raters, there was, of course, some variation. Rater 3 was the only rater who was not a language teaching professional. One wonders whether teachers

Table 4: A Summary of Descriptive Statistics

Variable	n	Mean	Std Dev.	Mini	Maxi	Range
TOEFL	25	502.48	34.03	423.00	577.00	154.00
LT	25	49.48	3.86	43.00	59.00	16.00
ST	25	51.28	4.74	42.00	64.00	22.00
RD	25	50.00	4.62	38.00	59.00	21.00
SA	25	86.08	14.59	60.00	116.00	56.00
LL	25	77.05	8.49	61.00	97.70	36.70
LL1	25	81.60	10.03	65.00	101.00	36.00
LL2	25	84.36	11.14	63.00	110.00	47.00
LL3	25	65.20	8.98	47.00	84.00	37.00
OPDCT	25	92.48	6.70	77.83	110.90	33.07
OPDCT1	25	91.50	7.95	74.00	107.00	33.00
OPDCT2	25	95.11	7.88	75.00	107.00	32.00
OPDCT3	25	90.84	12.68	76.00	139.90	63.90
MCDCT	25	70.00	14.43	30.00	95.00	65.00
RPSA	25	78.88	14.31	61.00	111.00	50.00
RP	25	78.88	10.53	61.00	102.00	41.00
R1	25	78.60	11.28	60.00	104.00	44.00
R2	25	76.16	8.79	59.00	91.00	32.00
R3	25	81.88	13.66	62.00	112.00	50.00

(LT = Listening; ST = Structure; RD = Reading; SA = Self-Assessment; LL = Average of the three raters' scores for the test; LL1–LL3 = Raters' individual LL scores; OPDCT = Average of the three raters' scores for the Open Discourse Completion Test; OPDCT1–OPDCT3 = Raters' individual OPDCT scores; MCDCT = Multiple-choice Discourse Completion Test; RPSA = Role-play Self Assessment; RP = Average of the three raters' scores for the Role Play test; and R1–R3 = Raters' individual RP scores)

are considerably more tolerant of participants' efforts at appropriateness than non-teachers. Without other non-teacher raters, however, it is difficult to draw such a firm conclusion.

Similarly for the RP test, the rater with the lowest mean, Rater 2, was British, whereas the other two raters were Americans. One wonders whether the British rater tended to rate students lower due to higher expectations of what constitutes appropriate language use, having come from a country noted for its emphasis on politeness. Again, it is impossible to draw such a conclusion with just one rater, but it would be interesting to experiment with a large pool of raters to see if there is quantifiable variation in the way raters from different English-speaking countries (and/or cultural backgrounds) rate students.

Reliability

Internal consistency reliability

Internal consistency[2] reliability was computed by first using the split-half method to determine the correlation between odd- and even-numbered items in the test. The half-test correlation was then adjusted using the Spearman-Brown Prophecy formula to estimate full-test reliability. Table 5 shows the estimated full-test reliability of each of the six tests. The two tests in which students assessed themselves, the SA and RPSA tests, showed particularly high estimates of internal consistency, followed by the LL and RP tests. Both of the discourse completion tests, especially the MCDCT, had considerably less internal consistency.

Table 5: Adjusted Split-Half Internal-Consistency Estimates

SA	LL	OPDCT	MCDCT	RPSA	RP
.9567	.9260	.6711	.5612	.9304	.8636

Standard Error of Measurement

The Standard Error of Measurement (SEM)[3] was computed using the standard deviation estimates from Table 4 and the adjusted split-half values from Table 5. Table 6 shows the SEM for the six tests. As can be seen, the LL test yielded the smallest SEM at 2.3, whereas the MCDCT clearly had the highest at 9.55. The others had respectable estimates of SEM in the 3.0 range.

Table 6: Standard Error of Measurement

	SA	LL	OPDCT	MCDCT	RPSA	RP
SEM:	3.03	2.30	3.84	9.55	3.77	3.88

Validity

Content validity

Since there is no statistical measure of content validity, either the testers themselves, their colleagues, or panels of experts determine the "representativeness and comprehensiveness" of the tests (Hatch & Lazaraton, 1991, p. 540). To ensure content validity, Hudson et al. have created a framework in which the speech acts of requests, apologies, and refusals are systematically matched with the variables of relative power, social distance and degree of imposition. According to Hudson et al., "[t]he designation of these in this way allows an examination of the interaction between sociopragmatic variables and particular speech act realizations. Additionally, this framework allows an examination of each particular variable within each speech act" (1992, p. 16). Furthermore, the role-play situations involve a wide and fairly representative sampling of real-life contexts: interacting with a mechanic at a garage, with a clerk at a store, with a superior in the workplace, with a housemate in a shared house, etc.

Criterion-related validity

Criterion-related validity involves comparing the results of the test or tests being evaluated with some other established measure of proficiency (Brown, 1996, p. 247). We chose the students' TOEFL scores for comparative purposes for a variety of reasons: 1) we had ready access to these students' TOEFL scores since they had taken an institutionally-administered TOEFL examination several months earlier upon entrance into our university; 2) students' TOEFL scores have proven reasonably effective for placement purposes within our own English language program; and 3) TOEFL scores are widely used and accepted as a measure of a student's overall English language proficiency. First, correlation coefficients were determined between the students' TOEFL subtest scores of Listening (LT), Structure (ST), and Reading (RD), and the tests of this study—SA, LL, OPDCT, MCDCT, RPSA, and RP.

These correlations were then squared to find the *coefficient of determination.*[4] The coefficient of determination ascertains the amount of overlapping variance between the tests, in effect revealing which correlations are meaningful. The results of squaring the above values to yield the percentage of overlapping variance between the tests are in Table 7. As can be seen, the only significant amount of overlapping variance is within each set of tests. The greatest amount of overlap is between the ST and RD tests at .359, an overlap of approximately 36%. The next greatest amount of overlap is between the production-based pragmatic tests, especially between that of the LL and OPDCT at approximately 29%, and between the LL and the RP also at nearly 29%. Further overlap can be found between the two self-assessment tests, the SA and RPSA, at approximately 22%. Within each set of tests, then, there is some *meaningful* overlapping variance between certain tests, but essentially no overlapping variance between the set of tests designed by Hudson et al. and the TOEFL subtests. It seems quite clear that these two sets of tests are measuring something very different from one another.

Table 7: Squared Correlation Values to Determine Overlapping Variance

	LT	ST	RD	SA	LL	OPDCT	MCDCT	RPSA	RP
LT	1.000								
ST	.169	1.000							
RD	.014	.359**	1.000						
SA	.000	.002	.003	1.000					
LL	.097	.050	.014	.022	1.000				
OPDCT	.022	.007	.018	.008	.287*	1.000			
MCDCT	.013	.004	.003	.110	.028	.051	1.000		
RPSA	.000	.046	.009	.217*	.001	.114	.050	1.000	
RP	.019	.017	.100	.000	.285*	.156	.001	.050	1.000

*p < .01
**p < .001

Construct validity

Principal component analysis (PCA): A principal component analysis[5] of the TOEFL subtests and the six tests of pragmatic competence by Hudson et al. determined that there are three factors with Eigen values of over 1.0. The largest of these, Factor 1, accounts for approximately 24% of the variance, followed by Factor 2 accounting for approximately 22%, and Factor 3 at approximately 19%. Cumulatively, these factors account for approximately 65% of the variance.

Factor analysis: A factor analysis[6] using a varimax rotated factor matrix was then run in order to determine whether there was a pattern to the factor loadings. As shown below in Table 8, results after a varimax rotation of these factors show a clear pattern of factor loading by test type, with the highest load on three of the tests by Hudson et al., closely followed by the TOEFL subtests, and then by the two self-assessment tests. This strongly suggests that some sort of method effect is at work.

That is, each of these types of tests seem to have factors in common which are not shared by the other tests. What these factors are is not clear, but one can speculate. The LL, OPDCT, and RP tests are similar in that they all employed native speakers of English rating the students' actual production of English: spoken, written and in role-play situations, respectively. The TOEFL subtests share the qualities of being paper and pencil tests that draw upon the students' receptive processes and require as a response the recognition of right answers in a multiple choice format. The SA and RPSA tests both involve the participants evaluating themselves, which is a method quite the opposite from the MCDCT.

Table 8: Factor Analysis

	Factor 1	Factor 2	Factor 3
Insert			
LT	.209	**.635**	.114
ST	.082	**.905**	.076
RD	-.351	**.732**	-.163
LL	**.867**	.229	-.004
OPDCT	**.728**	-.177	-.327
RP	**.790**	.018	-.185
SA	.145	-.095	**.730**
RPSA	.033	.077	**.823**
MCDCT	.197	-.087	-.630

Differential groups: Another method for determining construct validity is through an analysis of differential groups.[7] The participants in this study, it may be recalled, were divided into three different groups based on the length of their overseas experience. Group 1 had spent little or no time overseas, Group 2 from 5–10 weeks, and Group 3 a year or more (Table 1). Since in these tests the construct is pragmatic competence, it would be expected that the group with the greatest amount of time overseas in English-speaking environments would have the greatest amount of pragmatic competence.

A multivariate analysis of variance (MANOVA) procedure showed that there were significant differences among these three groups in terms of their test results. Univariate follow-up statistics were then run to determine the extent to which each of the tests differentiate between these groups, as given in Table 9 below.

Table 9: Univariate Follow-up Statistic

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig of F
LT	18.898	339.341	9.449	15.424	.612	.551
ST	29.965	509.075	14.982	23.139	.647	.533
RD	66.408	445.591	33.204	20.254	1.639	.217
SA	515.098	4594.741	257.549	208.851	1.233	.311
RPSA	1191.190	3725.450	595.595	169.338	3.517	.047*
RP	1352.64	1310.443	676.320	59.565	11.354	.000**

*$p < .05$

**$p < .001$

As indicated, the univariate follow-up statistic showed p values below .05 for two of the tests, the RPSA and the RP. Since these two tests yielded values at the $p < .05$ level, the Scheffé post hoc test was conducted to determine the significance of paired differences. For the RPSA test, the Scheffé test showed no two pairs of groups were significantly different at the .05 level. However, Scheffé post hoc analysis of the variance of the RP test, which had yielded a particularly low p value of .0004, showed significant Scheffé paired differences with the mean scores of Group 3 substantially and significantly different from either those of Group 1 or Group 2, as can be seen in Table 10.

Table 10: Scheffé Paired Differences Test for the RP Test

Group	Grp 2	Grp 1	Grp 3	Mean
Grp 2				74.3611
Grp 1				76.5417
Grp 3	*	*		93.4667

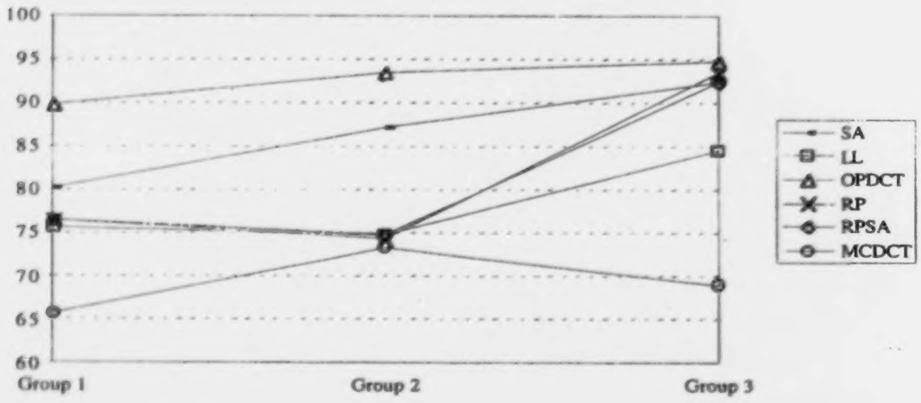
*$p < .0$

It is interesting to note that there is very little difference between Group 1, which had very little overseas experience, and Group 2, which had typically spent several weeks in English-intensive environments. In fact, Group 1 had a higher mean than that of Group 2, but this may have just been a random variation due to the relatively small number of participants in this study. That Group 3 had a much higher mean than either of the other two groups suggests that the development of pragmatic competence requires a substantial amount of time in the target culture.

Means comparison: A means comparison of the various tests offered further insight into the construct validity of the measures in this study (see Table 4 for all means). Among the TOEFL subtests there was very little differentiation between the three groups, and no clear patterns emerged from the data. The scores were very closely grouped by test for all three groups. The totals of the mean scores for each of the groups, in fact, were nearly the same, showing but a very slight increase by group: 150.36 for Group 1, 150.74 for Group 2, and 151.4 for Group 3.

With the tests of pragmatic competence, however, there was significantly more differentiation between the means scores of the groups. This can be seen in Figure 1. With the tests by Hudson et al., Group 3 clearly scored higher than the other two groups in all but the MCDCT test. This is particularly true of both the RP and the SA tests. The RP test, since it provides native speaker raters with a rich array of material on which to base their assessment, would be expected to provide the most accurate assessment of these students' pragmatic competence. It is interesting to note, however, that the RPSA scores are very nearly parallel with the RP scores, suggesting the students may be able to evaluate their own performance as well as the native speaker raters. The LL test also clearly differentiated the pragmatic competence of the Group 3 participants from those of Groups 1 and 2, while the SA and OPDCT

Figure 1: Means Comparision by Differential Groups—Pragmatic Tests



showed a small amount of differentiation. The MCDCT, however, was clearly out of synch with the other tests, and shows Group 3 to have less pragmatic competence than either of the other two groups.

A final point of interest is the disparity between the SA mean and the RPSA and RP means for Group 2, most of whom had recently returned from six-week overseas English-study experiences. On the SA test they seem to have been quite confident of their pragmatic competence as indicated by scores that, on average, were substantially higher than those for Group 1. After performing the role plays, however, Group 2 as a whole rated themselves a good bit downward, apparently feeling they had not performed nearly as well as they thought they could, which is confirmed by the very similar mean produced by the RP test. Group 1 also rated themselves downward after the RPSA, but not as much as Group 2 did. Group 3, on the other hand, appears to have been the only group that had a fairly clear idea of how well they could and did perform, as evidenced by very similar means for all three tests.

Test Practicality

The level of practicality of the multi-test framework—especially in terms of requirements related to time, number of personnel, and special equipment—varied greatly between the tests. Administering the OPDCT and MCDCT was relatively simple. Just a few minutes were required to hand out the tests and instruct students on how to complete the test at home. Taking the tests, however, did require quite a bit of time, especially the OPDCT. The SA test was also easy to administer. All could take it simultaneously, and it did not require much time nor any special equipment.

Administering the other tests was considerably more involved. For the LL, two cassette tape recorders were required; one for playing the situations, and the other for student responses. Additionally, the test needed to be conducted in a quiet room free from disturbances, and the participants needed to take the test individually. Some 10 minutes were required per student to set them up with the equipment and test. Of the six tests, the greatest amount of time and energy was required to administer the RPSA and RP tests. Although these two tests could be conducted concurrently (the data provided by performing the role plays could be used by the students to rate themselves as well as by the raters), performing a full set of role-plays required some 30 minutes per student. The RP test additionally required that the role plays be recorded on video tape so that these recordings could be distributed for evaluation by each of the raters.

Conclusions

With the exception of the OPDCT and MCDCT, the tests designed by Hudson et al. proved highly reliable and valid in assessing pragmatic competence when administered to Japanese university EFL students. The TOEFL subtest scores, by comparison, did not correlate with the pragmatic competence of the students. It would appear as well that the development of pragmatic competence requires fairly extended periods of time in the target culture for the realization of appreciable gains. A few weeks overseas in English-speaking immersion situations seems not to make much difference in learners' pragmatic competence—a year or more is required based on the results of this study. As for the practicality of administering and evaluating these tests, there was a great deal of variance. Of the four tests that proved both reliable and valid, only the SA test was easy to administer and evaluate, although the results were not as accurate as with those of the LL, RPSA, and RP tests.

One particular limitation of this study has to do with the representativeness of the participant group in terms of the variety of English speakers among native Japanese. The participants were all first-year university students with somewhat similar TOEFL scores, so lacked diversity in age, occupation, and linguistic ability. As suggested by Yamashita (1996), older learners involved in the work force would be more aware of the strict social conventions of Japanese society, making them perhaps more sensitive to sociolinguistic concerns in other languages as well. Native Japanese who use English in a service industry might also have a higher sensitivity to such concerns. Surely the linguistic ability of participants would have some influence on pragmatic competence as well, those with higher levels having a greater range of linguistic options available to them when attempting to be appropriate in a particular situation.

The potential directions of future research are many. As mentioned, having a wider range of participants would be desirable for determining the relationship between age and linguistic competence with pragmatic competence. As suggested earlier when discussing the variation in the ratings by the raters, it would be interesting to do rater comparisons between language teaching professionals and non-teachers to see if teachers have a higher acceptance of pragmatic incompetence than might non-teachers. Similarly, it would be interesting to compare raters from different native English speaking cultures to determine if there is, in fact, variation in standards of appropriateness by culture. Finally, there is the matter of examining the transcriptions of the student utterances in the role plays, for here lies a rich corpus of data for doing a qualitative analysis of these participants' pragmatic competence.

Acknowledgments

We would like to thank Sayoko Yamashita and Randy Thrasher of International Christian University, as well as J. D. Brown of the University of Hawai'i, for their assistance with both data analysis and proofreading of earlier drafts of this paper.

Ken Enochs teaches in the English Language Program at International Christian University, Tokyo.

Sonia Yoshitake-Strain, Ph. D., has taught at International Christian University and Seigakuin University, and is currently Japan Tutor for the Birmingham University MA in TESOL Program.

Notes

1. Making the adjustment for the three raters together involved converting the Pearson *r* values from Table 5 into Fisher *Z* coefficients using a Fisher *Z* transformation table (Guilford & Fruchter, 1978, p. 522). The Fisher *Z* coefficients were then averaged and converted back to Pearson *r* coefficients. These average figures were then adjusted to take into account the number of different raters using the Spearman-Brown Prophecy formula.
2. Internal consistency is an indirect way to estimate (without actually retesting) the consistency of a test. One common estimate of a test's internal consistency is to use the split-half method to first determine the correlation between odd and even numbered items in the test, and then adjust the half-test correlation using the Spearman-Brown Prophecy formula to estimate full-test reliability (Brown, 1996).
3. The standard error of measurement (SEM) is a statistic that uses both the standard deviation of a test and a correlation coefficient to "determine a band around a student's score within which that student's score would probably fall if the test were administered to him or her repeatedly" (Brown, 1996, p. 206).
4. The coefficient of determination, according to Brown (1996), shows the proportion of variance between the scores that is common to both, or the degree to which the two tests line up the students in the same order.
5. Principal component analysis involves determining "whether there are components that are shared in common by [several] tests and whether we can capture them in a meaningful way" (Hatch & Lazaraton, 1991, p. 490).
6. Factor analysis reduces a matrix of correlation coefficients to more manageable proportions, the result of which can be used to identify factors that the set of tests have in common (Alderson, Clapham & Wall, 1995, p. 289).
7. Analysis of differential groups determines the extent to which one group has more of the construct in question than another group (Brown, 1996, p. 240).

References

Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Beebe, L.M. & Takahashi, T. (1989). Sociolinguistic variation in face-threatening speech acts. In M. Eisenstein (Ed.) *The dynamic interlanguage* (pp. 199–218). New York: Plenum.

Beebe, L.M., Takahashi, T. & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R.C. Scarcella, E. Andersen & S.C. Krashen (Eds.), *On the development of communicative competence in a second language* (pp. 55–73). New York: Newbury House.

Brown, J.D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.

Canale, M. (1988). The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67–84.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,* 1, 1–47.

Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Oxford: Blackwell.

Enochs, K. and Yoshitake, S. (1996). Self assessment and role plays for evaluating appropriateness in speech act realizations. *ICU Language Research Bulletin, 11,* 57–76.

Guilford, J.P. & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.

Hudson, T., Detmer, E. & Brown, J.D. (1992). *A framework for testing cross-cultural pragmatics*. Honolulu: University of Hawaii Press.

Hudson, T., Detmer, E. & Brown, J.D. (1995). *Developing prototypic measures of cross-cultural pragmatics*. Honolulu: University of Hawaii Press.

Hymes, D. (1972). On communicative competence. In J.B. Pride & J. Holmes (Eds.) *Sociolinguistics*. Harmondsworth, Middlesex, UK: Penguin.

Mundby, J. (1978). *Communicative syllabus design: A sociolinguistic model for defining the context of purpose-specific language programmes*. Cambridge: Cambridge University Press.

Thomas, J. (1983) Cross-cultural pragmatic failure. *Applied Linguistics,* 4, 91-112.

TOEFL: Test and score data summary. (1995-1996). Princeton, NJ: Educational Testing Service.

Yamashita, S. (1996). *Comparing six cross-cultural pragmatics measures*. Unpublished doctoral dissertation, Temple University, Philadelphia.

Yoshitake-Strain, S. (1997). *Interlanguage competence of Japanese students of English: A multi-test framework evaluation*. Unpublished doctoral dissertation, Columbia Pacific University, California.

(Received May 18, 1998; revised December 11, 1998)

Appendix: Sample Items of the Six Tests

Self-assessment test (SA)

Situation 1:

You live in a large house. You hold the lease to the house and rent out the other rooms. You are in the room of one of your house-mates collecting the rent. (This house-mate moved in recently.) You reach to take the rent check when you accidentally knock over a small, empty vase on the desk. It doesn't break.

Rating: I think what I would say in this situation would be

very unsatisfactory	1	2	3	4	5	completely appropriate

Listening laboratory production test (LL)

Situation 2:

You are applying for a job in a company. You go into the office to turn in your application form to the manager. You talk to the manager for a few minutes. (The manager is impressed by your CV and wants to hire you.) When you move to give the manager your form, you accidentally knock over a vase on the desk and spill water over a pile of papers.

You say:

Open discourse completion test (OPDCT)

Situation 3:

You have recently moved to a new city and are looking for an apartment to rent. You are looking at a place now. You like it a lot (and talk to the manager for a few minutes). The landlord explains that you seem like a good person for the apartment, but that there are a few more people who are interested. The landlord says that you will be called next week and told if you have the place. However, you need the landlord to tell you within the next three days.

You say:

Multiple choice discourse completion test (MCDCT)

Situation 4:

You are a member of the local chapter of a national ski club. Every month the club goes on a ski trip. You are in a club meeting now helping to plan this month's trip. The club president is sitting next to you and asks to borrow a pen. You cannot lend your pen because you only have one and need it to take notes yourself.

a. Oh, sorry, it's my only one. Maybe John has an extra. Let me check.
b. I'm terribly sorry, this is the only one I have at the moment. Perhaps you might ask John?
c. No, I can't lend this pen. It's my only one.

Role-play self-assessment test (RPSA) & Role-play test (RP)

Situation 6:

Background 6a: You work in a small shop that repairs jewelry. You do not do the repairs yourself; a repairman comes in at night to do the repairs.

Now: A valued customer comes into the shop to pick up an antique watch that you know is to be a present. You need to go in the back room to get the watch, but the customer is standing in the way of the door.

Background 6b: The repairman has not repaired the watch yet, even though it was supposed to be ready.

Now: Go back out to the customer.

The interlocutor is the customer. He will:

- stand in front of the backroom door
- request watch and hand over the slip
- move after request to move
- accept that it is not ready, agree to come back tomorrow
- ask for change for the bus
- see you tomorrow

Note: Have no change in the till

Working at the Jewelry Repair Shop

1. **Request**	very unsatisfactory	1	2	3	4	5	completely appropriate
2. **Apology**	very unsatisfactory	1	2	3	4	5	completely appropriate
3. **Refusal**	very unsatisfactory	1	2	3	4	5	completely appropriate

Massive Input Through *Eiga Shosetsu*: A Pilot Study with Japanese Learners

Michael "Rube" Redfield
Osaka University of Economics

This paper introduces a new yet natural way of providing massive amounts of comprehensible input to learners of English as a Foreign Language (EFL). Learners watch popular contemporary movies in order to internalize the meanings presented in sounds and images. Then they read the accompanying *eiga shoshetsu* (movie tie-in novels) in order to convert meaning into the target language. In the pilot program using *eiga shoshetsu* described here, college learners made significant gains in listening, reading and vocabulary measures through reading the novels and seeing the movies.

本論では、英語学習者に多量の「理解可能な入力情報（comprehensible input）」を与えることができる、全く新しくかつ自然な外国語学習・教授法を紹介する。この学習・教授法は、まず学習者が興味を持てる映画を鑑賞して音声情報と映像情報に支えられた英語の意味を学習し、その後、その映画小説を読むことで、すでに学習した英語を確かな知識として定着させる試みである。本論で紹介するパイロット研究では、リスニング、読解、語彙のそれぞれの学習において、この学習・教授法の教育的・統計的効果が確認された。

It has been suggested that a major reason for the relative failure of the English educational system in Japan to produce more communicatively competent learners is lack of exposure to significant amounts of meaningful input in the target language (see Koike, 1991, for a discussion of the problems facing English education). My own research has shown that typical Japanese college EFL students usually cannot read English with proficiency (Redfield, 1992b, 1994a; 1994b; 1995), often do not have grammatical accuracy (Redfield, 1990, 1991a, 1991c, 1992a) or good listening skills (Redfield, 1991b), although they can learn to listen (Redfield & Campbell, 1996), and often do not improve significantly from one year to the next (Redfield, 1994c), even after spending up to 800 classroom hours studying EFL (Redfield, 1992b).

JALT Journal, Vol. 21, No. 1, May, 1999

Other researchers have suggested that EFL writing instruction may not necessarily improve learners' writing skills (Robb, Ross & Shortreed, 1986). As one way of addressing this problem, the following report introduces methodology for delivering massive amounts of authentic, thematically interesting, comprehensible input into the Japanese college curriculum in order to provide students with more exposure to meaning-focused use of English.

The Role of Comprehensible Input in Promoting Language Acquisition

A number of language acquisition specialists have advocated the use of what has come to be known as the Comprehension Approach (Nord, 1974, 1975, 1980, 1981; Redfield, 1991b). At the base of the approach lies the idea that comprehension is a requisite for learning. Simply phrased, if learners do not in some way or another understand the meaning of what they encounter in their learning environment, be it in written or oral form, then the learners do not learn. Regardless of whether one is inclined to support the strong version of the Interaction Hypothesis (Ellis, 1991; Long, 1981, 1983, 1985), asserting that comprehensible input leads directly to language acquisition (Krashen, 1981, 1982, 1985; Pienemann, 1984, 1989), or the weaker version of the hypothesis, that comprehensible input under certain restraints can, but does not necessarily, lead to acquisition (Ellis, 1986, 1988, 1990; Fotos, 1993; Fotos & Ellis, 1991; Schmidt, 1990, 1992; Sharwood Smith, 1981; White, 1987), both researchers and classroom practitioners would agree that without comprehensible input no meaningful language acquisition is likely to take place. A corollary is that more input is probably better for learning than less input. The amount of comprehensible input matters. Once these fundamental ideas behind foreign language acquisition are understood and accepted, it then becomes a matter of applying this knowledge to classroom practice.

If what the leading researchers such as Long, Krashen and Ellis suggest is correct—that learners need massive amounts of comprehensible input in order to acquire foreign languages and since such massive input is not automatically available in the English as a foreign language environment—then we as classroom instructors should attempt to provide such input. The study described below presents one such effort.

Extensive Reading to Provide Meaningful Input

Krashen claims that one of the most effective ways to provide input is through reading (1982, 1985, 1989). Mason and Krashen (1997) present evidence from Japan suggesting that the use of graded readers in an extensive reading program can improve reading scores. Today most scholars recommend using authentic reading materials, and I have a related suggestion. Students should read what is known in Japan as "*eiga shosetsu*," the script-based English-language novel about an English-language movie that is published at the same time as the movie so that viewers can preview the movie or read about the theme in more detail after viewing it. Unlike novels upon which movies are based, where the two different versions, print and celluloid, clash more often than not, *eiga shosetsu* have the advantage of following the plot accurately right down to the dialogue. Unlike screenplays or tape scripts, *eiga shosetsu* have narrative and descriptions as well as dialogue. Making no pretensions towards literature, they are eminently easy to read. A particularly significant point is that if the EFL learner sees the film first, she/he already has absorbed the meaning of the story. As a previewing activity *eiga shosetsu* are equally as good. Here, the learner reads the book first, which facilitates processing the meaning of what is heard during the movie. *Eiga shosetsu* are popular with college-aged learners since they represent authentic use of the target language and are relatively easy to read. When read rapidly for enjoyment, they potentially provide massive meaning-focused comprehensible input. The trick, of course, is to get the learner to read them, and then to provide objective evidence that reading *eiga shosetsu* actually helps learners acquire English. That is what the present study attempts to provide.

Research Focus of the *Eiga Shosetsu* Pilot Program

It is suggested that the following positive results will be observed after Japanese college EFL learners are exposed to the massive amounts of meaning focused input involved in watching six English-language movies and reading seven English-language *eiga shosetsu* about movies they have watched.

Research Hypotheses

1. The learners will receive significantly higher scores on a reading post-test than they did on a reading pre-test.
2. The learners will receive significantly higher scores on a listening post-test than they did on a listening pre-test.
3. The learners will receive significantly higher scores on a vocabulary post-test than they did on a vocabulary pre-test.

Method

Participants

The 28 participants in this study were drawn from an intact group of 36 students taking an English composition class at a private Japanese university. The majority were English majors retaking the class as a required course after having failed it the previous year. Several English majors were taking the course for a third time. There were also education majors, a group of French majors, and a graduate student in literature taking the course as an elective. All of the students were upperclassmen (or above), meaning that they had had a minimum of eight years of formal English instruction, many a good bit more than the minimum. Their ability levels ranged from false beginner through elementary to intermediate, with two fairly advanced learners also taking part. One of these advanced learners had graduated from an international school in India, and the other had studied two years in San Francisco after graduating from a Japanese junior college. In other words, this was a very mixed group.

Procedures

The twenty-four week Japanese university school year was divided into six four-week sessions. Pre and post -reading, listening and vocabulary tests were administered to all students at the beginning and end of the six-session program. In the initial week of each session, the learners were shown the first part of a contemporary popular film. In the second week, the original film was viewed until its conclusion. In the third week the students were instructed to silently read the *eiga shosetsu* corresponding to that particular film. Students who did not have the correct book with them were allowed to read other material in English, often *eiga shosetsu* that they had not yet finished. The fourth session was devoted to writing a film review on the movie in question. Students were thus asked to read one *eiga shosetsu* per month as homework.

The movies chosen for viewing were *Dead Poets' Society*, *My Girl*, *The War*, *Braveheart*, *The Net*, and *The Assassins*. The students were also required to read a novel of their choice as summer vacation homework (most, but not all, choosing other unrelated *eiga shosetsu*). Weekly homework journals were also kept, assigned by the instructor on themes related to the movies. Except for written comments in the students' journals, there was no overt language instruction in the class.

In order to encourage students to complete the assignments, each student was asked how many pages he had read on the current *eiga* *shosetsu* each week when the class role was called. In order to demonstrate that the instructor believed that massive comprehensible input is necessary for second language acquisition to take place, during the silent reading sessions the instructor read a novel in Spanish. Although many of the learners probably did not finish all seven novels (the six assigned during the school year, and the seventh read as summer homework), they read at least parts of all of them, as witnessed by the instructor during the silent reading sessions. Even the least diligent members of the class averaged at least fifty pages read per novel, for a minimum total of 350 pages. The most diligent students read all seven novels, for an estimated total of over 2,000 pages. And all learners saw the six films for an additional 10-12 hours of aural input. Furthermore, many of the learners reported viewing the films at home a second time for more listening practice.

In summary, the *Eiga Shosetsu* Pilot Program required the students to watch six contemporary films, read seven movie tie-in novels, and write seven formal film/book reviews. The reading and viewing activities were designed to furnish massive comprehensible input.

Pre and Post-Testing

Three tests, a reading test, a listening test, and a vocabulary test, were administered on the first day of class in April, 1996 and again on the last day of the academic year in January, 1997. The results were scored, tabulated, and statistically analyzed using the *StatView* (1988), *JMP* (1994), *DataDesk* (1995), and *Statistica* (1994) statistical packages for the Macintosh computer. Out of an original class of 36, 28 learners took both the pre- and post- tests in two areas, and 26 took both tests in the third area. Students who only took the tests during a single administration were eliminated from the study. The tests are described in detail below.

Reading Test

The Scholastic Research Associates Reading Laboratory (SRA) is a well-known reading program used in the US to improve learners' reading abilities. The accompanying SRA Placement Test measures American grade school children's reading skills. It consists of two reading passages followed by five and nine (for a total of 14) reading comprehension items respectively. Each passage is timed, with students having exactly three minutes to complete reading the passage and to answer the multiple choice questions accompanying each reading. The same version of the SRA Placement Test was administered as both the pre-test and the post-test. The test is easy to administer, score, and interpret. It also has proven reliability with American learners.

Listening Test

The Campbell Listening Test (CLT) was developed by Professor Peter D. Campbell (Campbell & Redfield, 1996) to measure Japanese students' listening abilities in English. The test consists of 30 multiple choice items, based on grammar and vocabulary found in the Mombusho's school curriculum. The test is administered by playing an audio cassette containing instructions in both English and Japanese and the 30-item sentences, read by a female native speaker of "mid-Pacific" English. Students have an answer sheet only. Administration of the test takes approximately 25 minutes. The test was normed with Japanese college students drawn from the same population as those involved in the present study, and has a reported reliability of .8429 (Campbell & Redfield, 1996).

Vocabulary Test

The vocabulary level test was a modified version of Nation's Academic Vocabulary Test (AVT) (Nation, 1990). It consists of 18 items from each of five levels of a word count list, for a total of 90 items. The items were randomly selected from the 2,000, 3,000, 5,000, 10,000 and university word level lists. Participants had to match sets of three definitions from a column on the right with six words in the column on the left. There were six sets of three items each for each of the levels, for a total of 90 items. Learners were allowed 30 minutes to complete the vocabulary test. Although not normed with Japanese college learners, the test is purported to be highly reliable.

Statistical Analysis

For each test, the pre and post-test scores were combined to check the distribution, with a Shapiro-Wilk W test (Hatch & Lazaraton, 1991) performed to determine if the distribution was normal. Descriptive statistics were then calculated and differences between the pre and post-test scores were analyzed to determine whether they were significant using a paired one-tailed t-test. However, because there were only 26 participants (t-tests should be used when there are 30 or more participants), the non-parametric Wilcoxon Matched Pairs procedure (Hatch & Lazaraton, 1991) was also performed. The alpha level for statistical significance was set at the .05 level, usual for studies in the field.

Results

Reading

As described above, the pre and post-test SRA scores were combined to check the distribution. A Shapiro-Wilk W test was performed to determine if the distribution was normal. It was, barely ($W = 0.9512$, $p < 0.0584$). Descriptive statistics were then calculated and differences between the pre and post-test performances were observed (Table 1). A paired t-test was performed to determine the significance of the difference between the pre and post-test scores ($t = 7.759$, $p < .0001$). The post-test scores were significantly higher than the pre-test scores. Thus the learners improved significantly over the course of the year.

Table 1: Reading Test Descriptive Statistics

	Number	Mean	Std. Dev.	Std. Err.
Pre-test	26	6.577	1.579	.31
Post-test	26	9.769	1.966	.386

As mentioned, since there were only twenty-six subjects taking this test, the non-parametric Wilcoxon Matched Pairs procedure was also performed ($z = -4.197$, $p = .0001$). This test also indicated that the students scored significantly higher on the post-test than on the pre-test. The first hypothesis regarding significant reading gains was therefore confirmed.

Listening

Again, the pre and post-test CLT scores were initially combined to check the distribution. A Shapiro-Wilk W test was then performed to determine if the distribution was normal. It was ($W = 0.9637$, $p < 0.1813$). Descriptive statistics were calculated (Table 2) and a paired t-test was performed ($t = -2.195$, $p < .0184$). The post-test scores were again significantly higher than the pre-test scores. It is therefore suggested that the *eiga shosetsu* program led to progress in listening.

Table 2: Listening Test Descriptive Statistics

	Number	Mean	Std. Dev.	Std. Err.
Pre-test	28	16.786	5.1521	.974
Post-test	28	18.464	4.67	.883

Again, because of the limited number of students, the Wilcoxon Matched Pairs procedure was also performed ($z = 1.991$, $p < .0465$). Here as well significant gains were observed. The second hypothesis was therefore confirmed.

Vocabulary

Following the same procedures, the pre and post-test vocabulary scores were combined to check the distribution. A Shapiro-Wilk W test was then performed ($W = 0.9765$, $p < 0.5575$), indicating that the distribution was normal. Descriptive statistics were calculated (Table 3) and a paired t-test performed ($t = -2.469$, $p < .0101$). Again, the post-test scores were significantly higher than the pre-test scores, indicating that the learners had improved significantly over the course of the year. Thus, the *eiga shosetsu* program led to significant progress in vocabulary acquisition. However, once again because there were only 28 participants, the non-parametric Wilcoxon Matched Pairs procedure was also performed ($z = -2.362$, $p < .0182$). Here, as well, the students scored significantly higher on the post-test than on the pre-test, which, it is suggested, can be attributed to the *eiga shosetsu* program. The third hypothesis was therefore confirmed.

Table 3: Vocabulary Test Descriptive Statistics

	Number	Mean	Std. Dev.	Std. Err.
Pre-test	28	53.25	7.881	1.489
Post-test	28	56.003	8.792	1.661

Discussion

As indicated by the significant gain scores in reading, listening and vocabulary comprehension, the results of the *Eiga Shosetsu* Pilot Program were most satisfactory, especially the reading results. As measured by the SRA Placement Test, the participants improved an average of over 1.5 grades in reading skills over the course of a year, from roughly beginning third grade, second semester, to final fourth grade, second semester. This is impressive because it had taken the learners at least eight years to reach the third grade level in reading, and yet, after a single course, they were now almost at the fifth grade level. Massive pleasure reading of the seven *eiga shosetsu* is suggested to be the reason. To paraphrase Frank Smith, students learn to read by reading (Smith, 1982).

Although no formal student program evaluation was included in the pilot study, informal conversations and written journal entries indicate that the participants felt that it was easier to read at the end of the program than it had been at the beginning. When the students first took the SRA Placement Test, they had a difficult time, even though the class carefully went over a sample test before taking the actual exam. It appeared that these students had little experience of reading for meaning, especially under time constraints. At the end of the program, however, they easily completed the SRA Test.

There were also significant gains in listening ability. After watching six movies, reinforced through the subsequent reading of the movie tie-in book, these learners significantly improved their English listening skills, as measured by the Campbell Listening Test. Although the gains were not as dramatic as those evidenced in reading, these learners still improved over 5.5% over the course of the program. Massive input through twelve hours of movie viewing is suggested to have significantly improved the learners' listening scores since this was the primary listening activity of the course. All of this, it should be emphasized, was a result of massive input through pleasure viewing, and not a result of direct instruction.

The positive listening results reflect those reported in a recent paper by Redfield & Campbell (1996), who found that students taught through the medium of English showed significantly higher listening gains scores as measured by the CLT than did students instructed through the medium of Japanese, even when the major objective of the course was not the improvement of English listening skills.

Vocabulary recognition, which is closely related to reading (Day, Omura & Hiramatsu, 1991; Jenkins, Stein & Wysocki, 1984; Nagy, Anderson, & Herman, 1987; Krashen, 1982, 1989) also showed significant improvement over the course of the program, although to a lesser degree than reading and listening. As measured by Nation's Academic Vocabulary Test, the participants improved about 3% during the year. However, after reading up to seven novels, one might expect more substantial gains. Both the material read and the instrument chosen to measure vocabulary might have acted to limit the gains.

Eiga shosetsu are a type of easy reading. Although in no way can this be regarded as an objective measure, it took the researcher an average of less than an hour to finish reading each of the movie tie-in books used in the program. Although the books follow the movies down to the smallest detail (which is what makes them so attractive as teaching materials), they concentrate on simple narrative and dialogue. To this researcher, they fall somewhere between popular fiction and graded readers. As such, the vocabulary used is quite restricted. For pedagogical purposes, this is a plus, and one of the reasons behind developing the *Eiga shosetsu* Pilot Program in the first place. But reading works of a restricted vocabulary does not promote substantial gains on a vocabulary measure such as the AVT. This test measures words drawn from frequency count lists, and includes words at the 5,000, 10,000 and university vocabulary levels. It is doubtful that much vocabulary from the higher levels appears at all in movie tie-in literature, although this was not ascertained. However, it is suggested that a vocabulary test focusing on words from the 1,000, 2,000 or 3,000-word levels might have indicated larger gains.

A different way of measuring vocabulary knowledge might have resulted in more obvious vocabulary gains as well. Instead of having learners match definitions as a measure of vocabulary depth, one might, for example, follow Meara's suggestion (Meara & Buxton, 1987) and have learners simply indicate whenever they know a certain vocabulary word or not. Professor Campbell is working on just such as vocabulary measure, combining the limited vocabulary of the JACET Vocabulary List with the test procedures developed by Meara (Campbell, in preparation).

It is possible to suggest that the gains reported above resulted primarily from participation in the *Eiga Shosetsu* Pilot Program since all of the participants were upperclassmen who had taken all of their required English language courses. Thus, the composition class featuring the Pilot Program was the only English course the subjects were taking in the university. Certainly individual differences existed among participants and a number of outside factors could not be controlled; for example, several of the participants spent the summer of 1996 abroad and others might have been taking English classes at outside language schools. However, any gains registered by these participants did not arise as a result of work in other English classes because these learners were not enrolled in other English language classes.

Regarding suggestions for future research, the use of a control group consisting of a group of students from the same population studying in the traditional fashion without recourse to massive comprehensible input, would have been ideal. For the present pilot study, use of a control group was not possible. All efforts will be made to include a control group in the follow-up study.

Classroom Implications

Since the participants made significant gains by viewing, reading, and writing about movies, educators interested in achieving similar results in their own classes and programs should look to the different elements of the *Eiga Shosetsu* Pilot Program for ideas. Introducing a regular period of free pleasure reading into a typical 90-minute Japanese college class would be one obvious application. Showing contemporary films with required follow-up (such as movie reviews) is another. Initiating a reading homework program is a third, and having learners read a novel of their choice over the summer an obvious fourth. The key is to accept the theory behind the *Eiga Shosetsu* Pilot Program (i.e., that massive comprehensible input is necessary, if not sufficient, for second language acquisition to take place) and then develop appropriate course-specific applications of the theory.

Although the *Eiga Shosetsu* Pilot Program proved to be successful, it will necessarily be in need of constant modification. For example, because of the popularity and local availability of both movies and the corresponding *eiga shosetsu*, different movies will be introduced this year, with only *Dead Poets' Society* being retained from the previous program. Another change will be within the four-week sessions. Instead of playing the movie over the first two sessions, the first 90 minutes of the film will be played in the initial week only. The learners will then be required to rent the video themselves if they want to know the ending. There are two reasons for this change. First, if the learners rent the video in order to see the ending, they might be tempted, and certainly will be encouraged by the instructor, to watch the movie a second and third time, concentrating on listening closely to the English in an effort to improve their listening skills. It is hoped that they will not rely on reading Japanese subtitles.

The second reason has to do with a fundamental change in thinking about the use of class time. Rather than use class time watching the video and reading the book—activities which can be done outside of class—class time in the second administration of the program will be devoted to what can be done best in a social setting—interactive speaking and listening. Except for a brief 10-minute free reading warm-up period (introduced partially to check on the students' progress in reading the *eiga shosetsu* outside of class) at the start of each of the final three classes of the four-week session, class time during the last three weeks will be devoted to group and paired oral English practice. The second movie viewing, the silent reading periods, and the in-class review writing will all be moved outside of class. This, of course, is an experiment. Will the students actually do the work outside of class? The reason that movie viewing, reading and writing were initially structured as in-class activities was the lack of willingness on the part of the students to do homework. However, the thinking behind the change is that students need more than massive comprehensible input to master English; they also need time to interact with their peers and their instructor using English communicatively. This can best be done in a group setting and makes better use of class time. The question remains whether the learners will do the necessary outside work.

Conclusion

This paper describes the first administration of an experimental ELT program designed to provide massive comprehensible input to Japanese college students. Under the *Eiga Shosetsu* Pilot Program, twenty-eight university upperclassmen taking English composition class were asked to see six contemporary movies, read seven movie tie-in books, write seven movie/book reviews and keep a weekly journal. The learners took reading, listening, and vocabulary tests before and after finishing the nine-month program. On all three measures, the gains were statistically significant, suggesting that the *Eiga Shosetsu* Pilot Program was successful in raising participants scores on reading, listening, and vocabulary measures.

Future research includes modification of the program and this should also be studied to determine if the modifications were successful. Control groups should be included in further studies, and student evaluations of the program would be desirable. If the modified program also proves successful, it could be expanded to include learners from different faculties and institutions. Qualitative research might also be undertaken in order to see how the program affects individual learners. Student journals, think-aloud protocols, in-depth interviews, and ethnographic observations all come to mind. Finally, if the program consistently results in significant gains in reading, listening and vocabulary comprehension, then, with locally-mandated modifications, the program can be expanded to include learners from other cultures as well. All of these are deserving of further research.

Michael "Rube" Redfield teaches foreign languages, culture through sports, and computers at colleges in the Kansai area.

References

Campbell, P. Modifying the academic vocabulary test for Japanese learners. (In preparation)

Campbell, P. & Redfield, M. (1996). Reliability and the Campbell Listening Test: Part two, test-retest, split-half and Cronbach Alpha procedures. *The Osaka University of Economics Liberal Arts Review*, 14, 83-90.

DataDesk Version 5.0 . (1995). Data Descriptions, Inc., Ithaca, NY.

Day, R., Omura, C. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7, 541-551.

Ellis, R. (1986). *Understanding second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1988). *Classroom second language development*. New York: Prentice Hall.

Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Blackwell.

Ellis, R. (1991). The interaction hypothesis: A critical evaluation. In E. Sadatono (Ed.). *Language acquisition in the second/foreign language classroom* (pp. 179-211). Singapore: Regional English Language Centre.

Fotos, S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics*, 14, 386–407.

Fotos, S. & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly*, 25, 605–628.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.

Jenkins, J.R., Stein, M.L. & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21, 767-787.

JMP Version 3.1. (1994) SAS Institute Inc., Cary, NC.

Koike, I. (1991). English teaching system facing stiff test. *Yomiuri Daily News*, Jan. 17, 7.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Oxford: Pergamon Press.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73, 440-464.

Long, M.H. (1981). Input, interaction and second language acquisition. In H. Winitz (Ed.). *Native language and foreign language acquisition* (pp. 259-78). *Annals of the New York Academy of Science* 379.

Long, M. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, 17, 359–382.

Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.). *Input in second language acquisition* (pp. 377-393). Rowley, MA: Newbury House.

Mason, B. & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 4 (1), 1-12.

Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142-154.

Nagy, W., Anderson, R.C. & Herman, P.A. (1987). Learning word meanings from text during normal reading. *American Educational Research Journal*, 24, 237-270.

Nation, I.S.P. (1990). Teaching and learning vocabulary. *Language Testing*, 10, 27-40.

Nord, J.R. (1974). Why can't I just learn to listen? *American Foreign Language Teacher,* 4, 4-6.

Nord, J.R. (1975). The importance of listening. *The English Teacher's Magazine*, 24, 34-39.

Nord, J.R. (1980). Developing listening fluency before speaking: An alternative paradigm. *System*, 8, 1-22.

Nord, J.R. (1981). Three steps leading to listening fluency: A beginning. In H. Winitz (Ed.). *The comprehension approach to foreign language instruction*. Rowley, MA: Newbury House.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6, 186–214.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10, 52–79.

Redfield, M. (1990). University levels: A comparative study. *The Osaka University of Economics Liberal Arts Review,* 8, 24-34

Redfield, M. (1991a). University levels: A replication study. *Kei Dai Ronshu*, 42, 24-34.

Redfield, M. (1991b). Does the comprehension approach work? A pilot study in a college classroom. *Kei Dai Ronshu*, 42, 61-72.

Redfield, M. (1991c). Is university language teaching effective? A pilot study. *Kei Dai Ronshu*, 42(4), 57-63.

Redfield, M. (1992a). Evaluating college English: A replication. *Kei Dai Ronshu*, 43(2), 125-139.

Redfield, M. (1992b). The Lit. major: What they aren't learning. *Speech Communication Education* 5, 14-28.

Redfield, M. (1994a). We Japanese can all read . . . And we know they grammar, too. *Speech Communication Education*, 7, 19-28.

Redfield, M. (1994b). "Japanese can all read English..." A replica. *Kei Dai Ronshu*, 45(4), 105-110.

Redfield, M. (1994c). Are students really better now? An empirical study. *The Osaka University of Economics Liberal Arts Review*, 12, 27-32.

Redfield, M. (1995). We Japanese can read: Part three. Paper presented at the Communication Association of Japan International Conference, Sapporo University, Hokkaido. June.

Redfield, M. & Campbell, P. (1996). They can learn to listen: English as the medium of instruction in the college English class. *Osaka Kei Dai Ronshu*, 47 (2), 21-30.

Robb, T., Ross, S. & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality, *TESOL Quarterly*, 20, 83-93.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.

Schmidt, R. (1992). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206–226.

Sharwood Smith, M. (1981). Consciousness-raising and the second-language learner. *Applied Linguistics*, 2, 159–169.

Smith, F. (1982). *Understanding reading* (3rd ed.). New York: Holt, Rinehart & Winston.

Statistica Version 4.1. (1994). SoftStat, Tulsa, OK.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second language competence. *Applied Linguistics*, 8, 95–110.

StatView Version 1.03. (1988). Abacus Concepts, Berkeley, CA.

(Received March 13, 1998; revised July 23, 1998)

Influence of Personality, L2 Proficiency and Attitudes on Japanese Adolescents' Intercultural Adjustment

Tomoko Yashima
Kansai University

This research examines whether individual variables, including L2 proficiency and extroversion, affect the intercultural adjustment process of adolescent Japanese sojourners. A questionnaire was administered to 139 high school students studying in the United States for one year and to their host families. Multiple regression analyses were conducted with self-rated and host-rated measures of adjustment as dependent variables. Independent or predictor variables were standardized English test scores, extroversion scores as measured by a personality type indicator, and several variables taken from a pre-departure questionnaire. The results showed that extroversion was a predictor of almost all self-rated measures of adjustment, including satisfaction with friendship with Americans, relationships with the host family and school work. English proficiency was a predictor of host-rated adjustment. A stronger international interest and a less Japanese-centered outlook led to better academic adjustment and the participants' overseas experience was shown to positively affect host-rated adjustment measures.

本研究では、第二言語能力と外向的傾向を含むいくつかの個人要因が、日本人高校生の異文化適応に影響を与えるかどうかを明らかにすることを目的とする。調査対象は一年間アメリカの家庭にホームステイしながら、現地の高校に通う日本人高校生139名と、そのホストファミリーである。出発前に行った英語標準テストの成績、性格タイプインディケーターにより測定した外向的傾向、及び、質問調査から得たいくつかの変数を独立変数とし、質問紙郵送法による自己評価、ホスト評価の適応を依存変数として重回帰分析を行った。その結果、外向性傾向は、アメリカ人との交友やホストとの関係に対する満足度など、ほぼすべての自己評価の適応指標を予測できた。一方英語力はホスト 評価の適応を予測することができた。また、国際的関心をもち、日本中心性が弱いほど学校授業に対する満足度が高いこと、過去の海外滞在経験が適応に良い影響を与えることも示された。

JALT Journal, Vol. 21, No. 1, May, 1999

Research on intercultural communication has attempted to identify individual qualities and situational factors that facilitate adjustment to a new culture. A number of interpersonal communication skills have been isolated as universal qualities which lead to successful interaction with people in different cultures, e.g., role behavior flexibility, empathy, ability to display respect, tolerance for ambiguity, mindfulness and ability to reduce anxiety (Ruben, 1976; Gudykunst, Wiseman & Hammer, 1977; Hammer, Gudykunst & Wiseman, 1978; Brislin, 1981; Hawes & Kealey, 1981; Gudykunst, 1991; Kim, 1991).

Considering people's movements between cultures, however, it is clear that conditions vary greatly with regard to parameters such as the sojourners' mother culture and host culture (and the cultural distance between them), the purpose and length of the sojourn, the sociopolitical and economic conditions of the host country, and the ages and occupations of the sojourners. As these differences are likely to affect the adjustment process to varying degrees, a careful examination of individual sojourn cases to identify culture-specific, situation-specific problems is necessary.

Researchers have identified a number of difficulties that Japanese sojourners[1] face during their travels abroad. Some early studies claim that Japanese suffer maladjustment (Inamura, 1980) or culture shock to a greater extent than do people from other countries (Nakane, 1972). Ebuchi (1986) studied Japanese sojourners in Southeast Asian countries and reported a common interactional pattern of spending time with other Japanese nationals so as to avoid contact with members of the host culture. He calls this "adjustment through avoidance" as opposed to adjustment through interaction. However, in a fairly complete review of prior research on Japanese sojourners overseas, Okazaki-Luff (1991) argues that the claim that Japanese suffer more adjustment problems than other nationals has no empirical evidence. She concludes her survey by stating that the difficulties discussed in earlier research were often related to a lack of communicative competence in the host nation's language and culturally-based communication styles.

Communication Styles

Many researchers have discussed characteristics of Japanese communication styles by contrasting Japanese cultural values with those of the US, using key concepts such as independence/dependence, individualism/collectivism, and heterogeneity/homogeneity. Some show specific Japanese communication behaviors which are likely to hinder effective communication with non-Japanese (e.g., Ishii, 1984; Kawabata, 1987;

Moyer, 1987; Kume, 1989; Tanaka, 1991; Tezuka, 1992). According to Ishii (1984), in order to maintain harmony, verbal expression is often subdued in the Japanese culture, and ambiguity and vagueness are preferred over direct and clear cut expressions of one's opinion. He says that the communicator unconsciously "simplifies explanations rather than elaborates on them, and expects the other person to sense what is left unsaid" (p. 55). Hall (1976) analyzed this characteristic of Japanese communication in terms of the concept of high and low-context cultures. In a high-context culture, of which Japan is a typical example, most of the information is either in the physical context or internalized within the person, resulting in a tendency to depend less on language and other explicit codes for communication. Because of this, people from low-context cultures, who are less accustomed to having to guess what is not communicated explicitly, may have difficulty communicating smoothly with people from high-context cultures.

Cross-cultural empirical studies on communication styles suggest that Japanese are less inclined to talk (Geatz, Klopf & Ishii, 1990), are less assertive and responsive (Ishii, Thompson & Klopf, 1990), and demonstrate more reluctance for self-disclosure (Barnlund, 1975, 1989) than Americans. Further, in studies of psychological aspects of communication, Japanese were found to have more communication apprehension than Americans, Koreans, Chinese and Puerto Ricans (Klopf & Cambra, 1979; McCroskey, Fayer & Richmond, 1985) and were shown to be more introverted than British people (Iwawaki, Eysenck & Eysenck, 1977).

L2 Competence

In contrast to the amount of research that has focused on differences in communication styles in the study of intercultural communication and adjustment, not much emphasis has been placed on the sojourner's proficiency in the host country's language (Nishida, 1985; Uehara, 1992). Uehara attributes this to the fact that the bulk of earlier research in intercultural adjustment was conducted by British and North American researchers and it was assumed that the participants spoke English. Nishida (1985) argues likewise, "In most of the intercultural communication studies to date, researchers have not paid attention to the language spoken between the participants" (p. 249). Nishida calls attention to the fact that foreign language competence can be an important factor in situations where sojourners cannot communicate in their native/strongest language. In her study of 18 Japanese college students, listening and speaking skills in English were shown to correlate negatively with the culture shock they experienced during a four-week sojourn in America.

In one model of intercultural communication competence, foreign language proficiency is regarded as an aspect of "behavioral flexibility" (Gudykunst, 1991). Gudykunst states that "some attempt at using the local language is necessary to indicate an interest in the people and/or culture" (p. 123). For Japanese sojourners in America, where the host nationals for the most part are unlikely to speak Japanese, language is perceived as a major problem (Diggs & Murphy, 1991) or as one of the most important elements of international competence (Kawabata, Kume & Uehara, 1989). Studies of young Japanese show that local language development either precedes or coincides with the children's adjustment or acculturation process (Minoura, 1984; Farkas, 1983).

In preliminary studies conducted between 1989 and 1991[2] (Yashima & Viswat, 1991, 1993a) Japanese high school students sojourning in the United States for one year and their host families attributed the difficulty students faced to a lack of ability to communicate in English. Not only the students' actual competence in L2 but also psychological factors such as anxiety and lack of confidence in using the L2 were issues. The students also stated that in order to adjust to living in the United States it was essential to be outgoing, to have participatory behavioral patterns, and to have a willingness to open themselves up by talking with host nationals.

Thus, the students were faced with the difficult task of expressing themselves in a culture in which "openness," "a willingness to talk," and "a frank exchange of opinions " are valued, using a language in which they were not proficient (Yashima & Viswat, 1992, 1993a; Yashima & Tanaka, 1996).

Research Focus

The subjects of this study were Japanese high school students studying in the US. The research presented here examines whether or not objectively-assessed language competence and extroversion (sociability and talkativeness) can indeed predict Japanese sojourners' adjustment. Few studies have empirically examined the relationship between these factors (e.g., Nishida, 1985, mentioned above) and a causal relationship has not been clearly established. To address interpersonal aspects of adjustment, this study focused on those who have sojourned abroad long enough to overcome the initial period of culture shock and started to build relationships with members of the host culture.

Studies in the past (e.g., Iwao & Hagiwara, 1987; Diggs & Murphy, 1991) primarily relied on self-rated language skills as the basis for assessing language competence. However, while self-rated language skills may reflect some aspects of competency, they cannot be considered definitive. In addition, because adjustment studies on Japanese high school exchange students are scarce this researcher believes that the group deserves more attention, particularly since the number of adolescent participants in overseas study programs has increased in recent years. This group of subjects was also selected because of its relative homogeneity in terms of age, length and objective of sojourn, as well as similarities in their individual experiences (i.e., attending a local high school, homestaying with an American family).

Adjustment can be defined as a psychological state of comfort, satisfaction, and perceived acceptance by hosts (as in Brislin, 1981). As investigated here, adjustment also includes the aspect of interactional effectiveness as defined in terms of participation, social adjustment, or cross-cultural interaction, and transfer of skills (as in Ruben & Kealey, 1979).

In the case of high school sojourners, no tangible results such as transferring technical know-how, gaining a degree or concluding a business contract are expected. The purpose of the sojourn is to interact with Americans and improve speaking skills in English. Thus, forming good human relations with Americans is at the core of their adjustment process.

The Study

Research Questions

The following research questions were investigated:

1. Can the English language proficiency of a Japanese sojourner prior to departure (as tested by a standardized proficiency test) predict his/her adjustment in the United States?
2. Can the degree of extroversion tested by a personality indicator (as a holistic psychological indicator of outgoing behavioral tendencies, sociability and talkativeness) predict his/her adjustment in the United States?

In addition, attitudinal parameters related to the specific experience of "studying abroad" were examined as possible predictors of successful adjustment. They included motivational strength for interaction with Americans, motivation for language learning, former overseas experience, and international outlook.

Method

Participants

The participants were 139 Japanese high school students (94 females and 45 males) of 15 to 18 years of age, who lived with families and studied in America for one year.[3] In addition, their 139 host families participated in this study as respondents to a questionnaire. Prior to the students' departure, an orientation session was held in Japan, at which time part of the data was collected. One hundred and eighteen students (81 females and 37 males) attended this session. Sixty-one of the students who attended the orientation had previously been overseas, mostly for short trips of a few days to three weeks in duration.

Pre-departure Tests and Questionnaires

In the orientation session prior to departure, English tests, a series of questionnaires and a personality type test were administered, as described below.

Test of English

As a measure of English proficiency, the Secondary Level English Proficiency Test (SLEP) by ETS consisting of a 75-item listening comprehension section (SLEP 1) and a 75-item reading/grammar section (SLEP 2) was administered.[4] As an additional measure of proficiency, oral interviews were conducted with 45 out of 53 students who had been participants in the 1992-3 program. The interviews were rated by two TESOL specialists who were experienced in oral interview assessments. The students were rated on six aspects of oral proficiency.[5] The inter-rater correlation was .916. Moderately high correlations between the results of the SLEP and interview tests (Interview with SLEP 1: $r = .703$; Interview with SLEP 2: $r = .611$) suggest that SLEP 1 and 2 adequately measured the communicative English competence of Japanese high school students.

Pre-departure Questionnaire

The pre-departure questionnaire consisted of three sections written in Japanese: 1) a section asking for demographic information, 2) a motivation scale, and 3) a section designed to assess students' international outlook.

Motivation Scale

This consisted of 18 items designed to measure the student's motivation to study in America. The questionnaire was adapted from a previous study (Yashima & Viswat, 1993b) and used a 5-point Likert scale (1—"not at all important" to 5—"very important").

International Outlook

Nine items were adopted from the questionnaire used by Tanaka, Kohyama & Fujiwara (1991), using the same 4-point scale (1 - "I don't feel this way at all." to 4 - "I mostly feel this way"). This section was designed to assess the students' interest in and attitudes toward international affairs and foreign countries. These items are given in Table 2 in Results.

Personality Type Test

As a measure of personality type, a type indicator in Japanese, similar to the Myers-Briggs Type Indicator under development by *Jinji Sokutei Kenkyusho* in 1991 was used. This consisted of 105 questions of which 23 items were related to the extroversion/introversion dimension.[6] For each item, students were required to choose between two statements according to which better described their character.

Experience Abroad

The students were categorized into four groups depending on their length of stay in foreign countries: Group 1 had never been abroad; Group 2 had traveled abroad for a week or less; Group 3 had stayed abroad for three months or less but more than a week; and Group 4 had stayed abroad more than three months.

Measurement of Adjustment

Four months after their departure from Japan,[7] questionnaires were mailed to the students and their host families to assess the students' adjustment (see Appendix). The student questionnaire includes a measure of overall satisfaction, adjustment, and performance of social skills. The sections on adjustment and social skills were translated into English and then back-translated into Japanese by bilingual translators to ascertain the semantic and functional equivalence of the two sets of questionnaires. The English version was then sent to the host families. The items were selected based on the concept of adjustment discussed in an earlier section, referring to findings and information collected through preliminary studies conducted between 1988 and 1991. Two subsections of the questionnaire were analyzed for the purposes of the current study.

The Satisfaction Scale

This scale consisted of 20 items concerning various aspects of life in America such as "depth of friendship with Americans," "the amount of conversation with hosts," and "improvement of English." The students were asked to evaluate the degree of their satisfaction with each of these on a 5-point scale, from "1: dissatisfied" to "5: very much satisfied." A global measure of satisfaction is frequently used in sojourn studies (Uehara, 1986; Rohrlich & Martin, 1991). See Table 1 in Results.

Self-Rating of Overall Adjustment to Host Family and School

Overall adjustment to host family and school was rated on a 5-point scale from "1: not at all adjusted" to "5: very well adjusted." The host families were asked to rate the adjustment of the students they were hosting on an equivalent scale in the English questionnaire.

Of the 139 students, 116 returned the questionnaire. Among those, 17 had not taken the pre-departure tests. Therefore, 99 students completed both procedures. Among the 139 host families, 101 returned the questionnaire.

Analyses and Results

This report presents the statistical analyses and results together in three separate sections. First, the dependent variables or measures of adjustment are analyzed. Second, the independent variables or predictor variables are examined. Finally, the results of multiple regression analyses are reported. The *SPSS Statistics Package 6.1* for the Macintosh was used for the analyses that follow. Options used were Advanced Statistics and Professional Statistics.

Dependent Variables

Adjustment

Dependent variables were extracted from the adjustment questionnaires. The raw scores (1 - 5) of the self-ratings of overall adjustment and the host families' ratings of overall adjustment were used. To determine how items were clustered and to form categories for use as dependent variables, 20 items from the Satisfaction Scale were subjected to a factor analysis. The factor matrix appears in Table 1. Factor 1 receives fairly high loadings from six items pertaining to friendship, activities and conversation with Americans, and is labeled "satisfaction in friendships with Americans." Factor 2 loads heavily on five items concerning life with the host family and is labeled "satisfaction with host family." Five of the six items loading heavily on Factor 3 relate to school work, the other being "human development." This factor is therefore best labeled "satisfaction with school work." Factor 4 receives high loadings from three items, "school environment," "school atmosphere," "attitude of Americans in general towards the student," all of which seem to refer to the human and/or physical environment. This factor is labeled "satisfaction with environment." [8, 9]

One factor (international interest) derived from the questionnaire on International Outlook affected students' satisfaction with school work, and another factor (Japan-centeredness) almost attained the significance level. This means those who had stronger "international interest" and less "Japan-centeredness" were more likely to be satisfied with their school work.

Table 1: Factor Analysis of 20-item Satisfaction Scale
(Varimax Rotation, Principal-Component Analysis; N = 116)

Items in the questionnaire	Factors 1	2	3	4	Communality
Number of American friends	**.77**	.06	-.04	.41	.77
Depth of friendship with Americans	**.88**	-.00	-.00	.25	.83
Amount of conversation with American friends	**.86**	.10	-.02	.17	.78
Range of activities participated in with American friends	**.86**	-.01	-.02	.08	.74
Extra-curricular activities at school	**.55**	.05	.23	.02	.36
English development	**.58**	18	.32	-.27	.54
Closeness to host family	.12	**.81**	.27	.18	.77
Care by host family	.04	**.89**	.16	.18	.85
Food provided by family	-.02	**.88**	.11	.02	.78
Amount of conversation with host family	.20	**.82**	.18	.07	.75
Rooms and facilities at the host residence	-.01	**.70**	.21	.14	.56
Care by teachers	-.49	.16	**.76**	.43	.79
Teachers' teaching style	-.01	.14	**.79**	.26	.72
Content of classes	.23	.04	**.69**	.27	.60
Academic achievement	.04	.27	**.64**	-.02	.49
Participation in class	.04	.25	**.57**	.04	.39
Human development	.40	.29	**.50**	-.24	.55
School environment	.11	.15	.28	**.80**	.76
School atmosphere	.24	.16	.14	**.79**	.73
Attitude of Americans in general towards student	.28	.37	.21	**.61**	.63
Eigenvalues	6.67	3.27	1.91	1.55	
Percent of variance explained	33.3	16.4	9.5	7.8	

Factor 1: Satisfaction with friendships with Americans
Factor 2: Satisfaction with host families
Factor 3: Satisfaction with school work
Factor 4: Satisfaction with environment

Independent Variables

The independent variables in this study were: (1) the SLEP total score; (2) the score of extroversion by the type indicator; (3 and 4) the two factors from the International Outlook questionnaire; and (5 and 6) two items from the Motivation Scale, " to improve spoken English ability" and "interest in American people and culture." The International Outlook data will be presented first.

International Outlook

The nine items on International Outlook were scored along a 4-point scale. As a means of reducing the number of variables into fewer, more abstract categories to be used as predictor variables, a principal component factor analysis of these nine items was performed and yielded three factors as shown in Table 2. Factor 1 receives high loadings from four items: "interested in international events," "knowledgeable about Japanese culture," "have seldom been out of hometown (negative)" and "want to work in an area that will contribute to the development of the world" and is therefore labeled "international interest." Factor 2 loads heavily on three items that indicate patriotism and unwillingness to live outside of Japan and is labeled "Japan-centeredness." Factor 3 is defined by three items, "realize Japan's role and responsibility in the world," "familiar with life and manners in foreign countries," and "have awareness of and pride in being Japanese" and is therefore referred to as "awareness of being Japanese in the world."[10, 11]

Analysis of Variables

The other independent variables were analyzed as follows. The English test was scored using the supplied answer key, with raw scores rather than scaled scores used (150 points in total, Mean = 88.79, Standard deviation = 14.51, Reliability KR-21rk = .84). A total extroversion score was then calculated from the Personality Type Indicator results (Reliability KR-21rk = .79).

The independent variables selected were not strongly correlated with each other. Since International Outlook Factor 2 and Factor 3 showed a moderately high correlation ($r = .52$), Factor 3 was dropped from the analyses as it showed lower correlations with the dependent variables. As former overseas experience was considered to be categorical data, it was analyzed separately through ANOVA.

Multiple Regression Analysis

Multiple regression analyses using the stepwise method were conducted to examine whether English proficiency, extroversion and the other independent variables could predict eight measures of adjustment

Table 2: Factor Analysis of the Nine-item Questionnaire on International Outlook
(Varimax Rotation, Principal-Component Analysis; N = 116)

Items in the questionnaire	Factors 1	2	3	Communality
Interested in international events	**.75**	-.09	.02	.66
Knowledgeable about Japanese culture	**.66**	.38	.16	.63
Have seldom been out of hometown	**-.51**	-.03	.04	.81
Want to work in an area that will contribute to the development of the world	**.49**	-.29	.26	.55
Patriotic, have love for Japan	.14	**.86**	.07	.78
Do not want to live outside Japan	-.39	**.66**	-.07	.59
Realize Japan's role and responsibility in the world	.04	.10	**.86**	.76
Familiar with life and manners in foreign countries	-.00	-.08	**.74**	.73
Have awareness of and pride in being Japanese	.26	**.51**	**.53**	.63
Eigenvalues	2.26	1.65	1.18	
Percent of variance explained	25.1	18.4	13.1	

Factor 1: International interest
Factor 2: Japan-centeredness
Factor 3: Awareness of being Japanese in the world

assessed through the questionnaires.[12] The eight dependent variables were: (1-4) the four factors from the Satisfaction Scale shown in Table 1; (5) the students' self-evaluation of their adjustment with host families; (6) the students' self-evaluation of adjustment at school; (7) the host families' evaluation of the students' adjustment to the host family and (8) the host families' evaluation of the students' adjustment to school.

The results of the regression analyses are given in Table 3. As observed, the proportion of variance accounted for by the independent variables is not very great. However, the results indicate a significant contribution by some variables which is worth reporting. Extroversion was able to predict the students' satisfaction with friendships with Americans, their relationship with the host family, and their self-rated adjustment to the host family and to school. English proficiency, on the other hand, was the significant predictor of the host-rated adjustment of the students to their host families and school.

Neither item from the motivation scale could predict adjustment at the significance level of $p < .05$. Yet at three points the significance level

Table 3: Results of Stepwise Multiple Regression Analysis

Dependent Variables (Adjustment)	Independent Variables	Beta	F	R2	Adjusted R2***
Satisfaction with friendships with Americans	Extroversion	.32**	8.99**	.10	.09
	Culturally-oriented motivation		.21+		
Satisfaction with host family	Extroversion	.43**	18.57**	.19	.18
Satisfaction with school work	International interest	.30**	8.07**	.09	.08
	Japan-centeredness	-.22+			
	Culturally oriented motivation	.21+			
	Extroversion	.20+			
Satisfaction with environment	-		-		
Self-rated adjustment: Family	Extroversion	.24**	4.75*	.06	.04
Self-rated adjustment: School	Extroversion	.43**	18.47**	.19	.18
Host-rated adjustment: Family	English proficiency	.35**	8.93**	.13	.11
Host-rated adjustment: School	English proficiency	.31*	6.46*	.10	.08
	English-oriented motivation	.22+			

*p < .05
**p <. 01
+ p <. 1
***R2 is a coefficient of determination with a possible value between 0 and1. The closer R2 is to 1, the fitter the model. However, since R2 increases as the number of predictor variables is increased, R2 must be adjusted (Ishimura, 1992).

was nearly attained. Those who had a stronger interest in American people and culture before departure displayed a tendency towards being more satisfied with their relationships with American friends and school work, and those who had stronger motivation to study English tended to be rated higher by the hosts.

ANOVA revealed that host-rated adjustment to host families was significantly affected by group difference as shown in Table 4.[13]

Table 4: Result of ANOVA Investigating the Influence of Overseas Experience on Adjustment

D.F.	Sum of Squares between groups	Sum of Squares within groups	F Ratio
2/70	14.11	127.83	3.87 ($p < .05$)

Tukey's Honestly Significant Difference tests[14] were conducted to see whether there was any significant difference between any pairs of groups (Table 5). The results indicate that Group 3 (students who had been abroad up to three months but more than a week) had a significantly higher adjustment rating from their host families than Group 1 (students who had never been abroad) and Group 2 (students who had traveled abroad for a week or less). There was no significant difference between Groups 1 and 2.

Table 5: Pair-wise Comparisons with Tukey-HSD Tests: Three Student Groups

	Group 1	Group 2
Group 2	.15	
Group 3	-.90*	-1.05*

* $p < .05$

Discussion

With regard to Research Question One, which asked if the English language proficiency of a Japanese sojourner prior to departure could predict his/her adjustment in the United States, it was found that English proficiency was a significant predictor of the host family's evaluation of the students' adjustment to school and to life with the host family, but it did not predict the students' perceptions of adjustment or sense of satisfaction. This probably indicates that accurate verbalization is important from the host families' perspective. Students who appear to have adjusted in the host families' eyes are likely to be those who are communicating well in English, i.e. accurately and effectively.

As for the second Research Question, which asked whether the student's degree of extroversion could predict his/her adjustment, extroversion was found to be a predictor of almost all the self-rated measures of adjustment, and was most strongly related to the interpersonal aspects of adjustment, i.e., satisfaction with American friends and host families. Extroverted individuals tend to be sociable, and are able to initiate interactions and talk comfortably with strangers. They usually find it easier to communicate their intentions/emotions through verbalization and explicit communication behaviors. These qualities might have helped the students build relationships and experience satisfaction in relationships with American people.

Why, then, didn't extroversion predict the host families' judgment of the students' adjustment? The host family is a given environment where host parents are expected to play the role of caregivers. The family members might try to talk to the students, inviting them into conversation as some host parents mentioned in the questionnaires, and thus may allow the students to play a more passive role in communication. As a result, therefore, efficiency of communication based on accurate listening comprehension most likely becomes more important than the number of interactions initiated by the students, the latter being related to extroversion.

On the other hand, extroversion probably becomes more critical in situations such as the school, where the student needs to initiate interactions to build relationships. In such settings students need to interact with the social environment, to lay the groundwork for communication by, for example, approaching a classmate in a friendly manner, greeting and initiating a conversation, or joining a group of classmates having lunch. Another explanation may be that extroverted individuals who are communicative and active feel satisfied with themselves but, due to a lack of linguistic competence, they may not be viewed as interactionally effective by the host family. Other-rated adjustment in the school situation by teachers or friends would clarify this point.

How do other individual parameters affect the students' adjustment? It was shown that students who had a higher interest in international affairs and were more open-minded tended to be more satisfied with school-work and were academically better adjusted than those who were more close-minded. Stronger culturally-oriented motivation (an interest in American people and culture) has a tendency to lead to higher satisfaction in friendships with Americans and school life.

Past overseas experiences, if longer than a week, also seemed to facilitate adjustment. Those who had stayed abroad from eight days through three months had significantly higher adjustment ratings from their host families than those who had had a week or less overseas experience.

Conclusions

The results of these statistical analyses confirmed what has been reported previously based on preliminary interviews and students' self-reports (Yashima & Viswat, 1991, 1992, 1993a & b). In earlier studies, social skills were identified that were suggested to facilitate students' adjustment (Yashima & Tanaka, 1996). They included skills related to initiating interaction, self-exposure, participation and avoiding ambiguity pertaining to such activities as: "find and talk about shared interests with someone such as about sports or music," "participate in school activities, including clubs and preparation for school events," "volunteer to help with household chores," and "express feelings of satisfaction and dissatisfaction openly rather than hiding them." Social skills are, by definition, observable and learnable skills which facilitate individuals' social adjustment. They deal with "everyday, common, even apparently trivial situations which nevertheless cause friction, misunderstanding and interpersonal hostility" (Furnham & Bochnar, 1986, p. 241). Social skills training developed in clinical psychology is often designed to help people overcome a lack of confidence in interpersonal communication, but is usually offered in participants' L1 (Aikawa & Tsumura, 1996). Thus, although social skills which may be of help to the sojourners have been identified, the students need to learn to perform them in English. To this end, a previous report proposed an intercultural training program combining English teaching and social skills training that could be included in a pre-departure orientation (Yashima & Tanaka, 1996).

The results of this research confirm the usefulness of employing such training as part of an intercultural orientation program. Although English conversation classes are usually conducted to prepare students for living in America, for the most part what is taught is English for general purposes. This may not be of immediate help to the students in starting rapport-building interactions with friends at school or host family members. Designing a custom-made intercultural training course by incorporating a necessary skill-building component in English teaching sessions may facilitate the students' adjustment. All students, both introverts and extroverts, can learn to develop a broader repertoire of behaviors which will help them to interact effectively with North Americans. Such training appears to be target culture-specific, yet by learning the communication style of another culture, it is likely that students will be able to apply some of the skills they acquired when they encounter a third or fourth culture.

Cross-cultural adjustment offers a significant learning experience. As a result of what students learn though their overseas experience, it is hoped that they will be more "mindful" of the communication process, will develop greater "behavioral flexibility," and will have "reduced anxiety" in intercultural interactions. These are vital elements in the universal model of intercultural communication competence proposed by Gudykunst (1991). If this is the case they will probably be better able to cope with differences such as age, gender, and cultural background within Japan. In-depth case studies of several students' adjustment processes throughout the year's experience would be a useful follow-up study to shed light on the role of English competence and social skills in the adjustment and culture learning process, as well as the changes taking place in their attitudes, behaviors, and intercultural/interpersonal communication competence.

Acknowledgements

The author thanks Dr. Tomoko Tanaka, Professor Linda Viswat, and the anonymous reviewers for their suggestions and valuable comments on earlier versions of this paper. This research was supported in part by the Japanese Ministry of Education's Grant-in-Aid for Scientific Research.

Tomoko Yashima is Associate Professor of English and Intercultural Communication at Kansai University.

Notes

1. The word, "sojourners" is used in this paper to refer to people who spend an extensive period of time in an overseas country.
2. In these studies (Yashima & Viswat, 1991, 1993a), 40-50 minute interviews were conducted with 11 students who had just returned from the US after participating in the same program as discussed in this study. Subsequently, questionnaires consisting mostly of open-ended questions were sent to 108 students and 55 host families.
3. Fifty-three of the students stayed in the United States from the summer of 1992 to the following summer, while 27 stayed there from 1993 to 1994, 29 from 1994 to 1995, and 27 from 1995 to 1996.
4. The Secondary Level English Test developed by Educational Testing Service, Princeton, NJ, is a test used by the Japanese organizer who coordinates an Academic Year In America Program which sends students to the United States. TOEFL, a better-known standard test, was not used in this study because it was deemed to be too difficult for the Japanese high school students to be a reliable and valid indicator of their language proficiency.
5. The six aspects are grammar, pronunciation, attitude (willingness to speak and eagerness to continue a conversation), amount of information conveyed, appropriateness and overall fluency.
6. This type indicator, based on the Myers-Briggs Type Indicator, is designed to

assess four dimensions of human personality, one of which is extroversion/ introversion. See Briggs-Myers & Myers, 1980.

7. Experience and research have shown that there are distinct stages in the adjustment process as shown in the W-shape hypothesis (Gullahorn & Gullahorn, 1983). Our preliminary investigation based on this theory showed that more than 70% of the students had overcome the initial stage of culture shock and felt adjusted after three months in the United States (Yashima & Viswat, 1992).
8. Cronbach's alpha reliability for each factor was calculated. Factor 1: α = .86, Factor 2: α = .90, Factor 3: α = .81, Factor 4: α = .82.
9. The procedure suggested by Koyano (1988) was followed to arrive at these factors. The labeling procedures employed in Dornyei (1990) and in Verhoeven (1991) were also used to name the factors.
10. The procedures explained in the previous note were used here.
11. Cronbach's alpha reliability for each factor was calculated. Factor 1: α = .50, Factor 2: α = .55, Factor 3: α = .61.
12. A multivariate analysis rather than repeated multiple regression analyses is recommended for future studies, as the latter assumes the presence of different independent variables.
13. There were only four students who fell into Group 4 (students who had stayed overseas longer than three months). Three of them had stayed abroad for more than five years and the others for one year. They were excluded from the ANOVA, because they were too few in number to form a group, yet were too different in the length of their sojourn to be merged into Group 3.
14. See p.190 of *SPSS 6.1 Base System User's Guide* for the detailed procedure.

References

Aikawa, A. & Tsumura, T. (1996) *Shakaiteki sukiru to taijinkankei* (Social skills and interpersonal relationships). Tokyo: Seishinshobo.

Barnlund, D.C. (1975). *Public and private self in Japan and the United States*. Tokyo: Simul Press.

Barnlund, D.C. (1989). *Communicative styles of Japanese and Americans*. Belmont, CA: Wadsworth Publishing.

Briggs-Myers, I. & Myers, P.B. (1980).*Gift differing*. Palo Alto, CA: Consulting Psychologists Press.

Brislin, R.W. (1981). *Cross-cultural encounters: Face-to-face interaction*. New York: Pergamon Press.

Diggs, N. & Murphy, B. (1991). Japanese adjustment to American communities: The case of the Japanese in the Dayton area. *International Journal of Intercultural Relations*, 15, 103-116.

Dornyei, Z. (1990). Conceptualizing motivation in foreign language learning. *Language Learning*, 40, 45-78.

Ebuchi, K. (1986). Ibunka tekio no mekanizumu (The mechanism of cross-cultural adaptation). *Kyoiku to Igaku*, 34, 4-11.

Farkas, J. (1983). *Japanese overseas children's American schooling experience: A* *study of cross-cultural transition*. Unpublished doctoral dissertation. Ohio State University, Columbus.

Furnham, A. & Bochnar, S. (1986). *Culture shock: Psychological reactions to unfamiliar environments*. London: Methuen.

Geatz, L., Klopf, D.W. & Ishii, S. (1990). *Predispositions toward verbal behavior of Japanese and Americans*. Paper presented at the Annual Convention of the Communication Association of Japan, June, Tokyo.

Gudykunst, W.B. (1991). *Bridging differences*. Newbury Park, CA: Sage Publications.

Gudykunst, W.B., Wiseman, R.L. & Hammer, M.R. (1977). Determinants of a sojourner's attitudinal satisfaction: A path model. In B. Ruben (Ed.) *Communication Yearbook*, 1, 415-425.

Gullahorn, J.T. & Gullahorn, J.E. (1983). An extension of the U-curve hypothesis. *Journal of Social Issues*, 19 (3), 33-47.

Hall, E.T. (1976). *Beyond culture*. Anchor Books.

Hammer, M.R., Gudykunst, W.B. & Wiseman, R.L. (1978). Dimensions of intercultural communication effectiveness: An exploratory study. *International Journal of Intercultural Relations*, 2, 382-393.

Hawes, F., & Kealey, D.J. (1981). An empirical study of Canadian technical assistance. *International Journal of Intercultural Relations*, 5, 239-258.

Inamura, H. (1980). *Nihonjin no kaigai futekio* (The maladjustment of Japanese overseas). Tokyo: NHK Books.

Ishii, S. (1984). *Enryo-sasshi* communication: A key to understanding Japanese interpersonal relations. *Cross Currents*, 11, 49-58.

Ishii, S., Thompson, C. & Klopf, D. (1990). A comparison of assertiveness/responsive construct between Japanese and Americans. *Otsuma Review*, 23, 63-71.

Ishimura, S. (1992). *Suguwakaru tahenryo kaiseki* (Easy guide to multivariate statistical analysis). Tokyo: Tokyo Shoseki.

Iwao, S. & Hagiwara, S. (1987). *Zainichi ryuugakusei no tainichi imeji* (The image of international students toward Japan, No. 8: Differences with length of stay and Japanese language competence). *Keiogijuku Daigaku Shinbun Kenkyusho Nenpo*, 29, 55-75.

Iwawaki, S., Eysenck, S.B.G. & Eysenck, H.J. (1977). Differences in personality between Japanese and English. *The Journal of Social Psychology*, 102, 27-33.

Kawabata, M. (1987). Prospects and structure of the studies in intercultural education. *Intercultural Education*, 1, 8-17.

Kawabata, M., Kume, T. & Uehara, A. (1989). An exploratory study of intercultural qualities, abilities and attitudes for the Japanese. *Bulletin of the Center for Education of Children Overseas, Tokyo Gakugei University*, 5, 93-95.

Kim, Y. (1991). Intercultural communication competence. In S. Ting-Toomey and F. Korzenny (Eds.). *Cross-cultural interpersonal communication*. Newbury Park, CA: Sage.

Klopf, D.W. & Cambra, R.E. (1979). Communication apprehension among college students in America, Australia, Japan, and Korea. *Journal of Psychology*, 102, 27-31.

Koyano, W. (1988). *Tabenryo kaiseki gaido* (Multivariate statistical analysis guide). Tokyo: Kawashima Shoten.

Kume, T. (1989). Reverse culture shock of the Japanese youth and education for effective intercultural communication. *Intercultural Education*, 3, 52-67.

McCroskey, J.C., Fayer, J.M. & Richmond, V.P. (1985). Don't speak to me in English: Communication apprehension in Puerto Rico. *Communication Quarterly*, 33, 185-192.

Minoura, Y. (1984). *Kodomo no ibunka taiken* (Cross-cultural experience of Japanese children in the United States). Tokyo: Shisakusha.

Moyer, Y. (1987). Factors of psychological stresses and aspects of counter measures: The cases of foreign students in Japan. *Intercultural Education*, 1, 81-97.

Nakane, C. (1972). *Tekio no joken* (The conditions for adjustment). Tokyo: Kodansha.

Nishida, H. (1985). Japanese intercultural communication competence and cross-cultural adjustment. *International Journal of Intercultural Relations*, 9, 247-269.

Okazaki-Luff, K. (1991). On the adjustment of Japanese sojourners: Beliefs, contentions, and empirical findings. *International Journal of Intercultural Relations*, 15, 85-102.

Rohrlich, B. & Martin, J.N. (1991). Host country and reentry adjustment of student sojourners. *International Journal of Intercultural Relations*, 15, 163-182.

Ruben, B.D. (1976). Assessing communication competency for intercultural adaptation. *Group and Organization Studies*, 1, 334-354.

Ruben, B.D. & Kealey, D.J. (1979). Behavioral assessment of communication competency and the prediction of cross-cultural adaptation. *International Journal of Intercultural Relations*, 3, 15-47.

Tanaka, T. (1991). Social skills for cross-cultural adjustment of international students in Japan. *Intercultural/Transcultural Education*, 5, 98-110.

Tanaka, T., Kohyama, T. & Fujiwara, T. (1991). A study on the social skills of Japanese students in short term overseas language seminar. *Memoirs of the Faculty of Integrated Arts and Sciences III Hiroshima University*, 15, 87-102.

Tezuka, C. (1992). *Awase* and *sunao* in Japanese communication and their implications for cross-cultural communication. *Keio Communication Review*, 14, 37-50.

Uehara, A. (1986). The nature of American student reentry adjustment and perception of the sojourn adjustment. *International Journal of Intercultural Relations*, 10, 415-438.

Uehara, A. (1992). *Gaikokujin ryuugakuseino nihongo joutatsuto tekiouni kansuru kisoteki kenkyuu* (A study on Japanese language competence and adjustment of international students in Japan). *A Report on Research 63510137* supported by Grant-in-Aid for Scientific Research.

Verhoeven, L.T. (1991). Predicting minority children's bilingual proficiency: Child, family and institutional factors. *Language Learning*, 41, 205-233.

Yashima, T. & Viswat, L. (1991). A study of Japanese high school students' intercultural experience—"It's not a dream country, but I love America." Change in image of Americans. *Human Communication Studies,* 19, 181-194.

Yashima, T. & Viswat, L. (1992). Sojourner adjustment: The role of a support group. *Bulletin of Heian College*, 22, 49-59.

Yashima, T. & Viswat, L. (1993a). An analysis of communication problems of Japanese high school students and their host families. *Human Communication Studies,* 21, 181-196.

Yashima, T. & Viswat, L. (1993b). English proficiency and intercultural adjustment of high school students studying in America. Paper presented at the 32nd Annual Convention of JACET, September, Tokyo.

Yashima, T. & Tanaka, T. (1996). English teaching for intercultural adjustment using social skill training techniques. *Intercultural Education*, 10, 150-166.

(Received January 27, 1998; revised October 9, 1998)

Appendix

Self-Rated Adjustment Scales

5　あなたは現在のホストファミリーにどの程度馴染んでいたと思いますか。５段階で評価してください。

1	2	3	4	5
全く馴染んでいない家族行事にも参加しない、客のようだ				家族の一員のように馴染んでいると思う

6　あなたは学校生活ににどの程度馴染んでいたと思いますか。５段階で評価してください。

1	2	3	4	5
友達もできず、学校が楽しくない				友達も大勢でき、学校の活動にも積極的に、参加し、とても楽しんでいる

Satisfaction scale

II あなたの留学中の生活について、次に示す項目のそれぞれにどの程度満足しているか評価してみてください。

5 非常に満足している
4 まあまあ満足している
3 ふつう
2 あまり満足していない
1 非常に不満足だ

1	学校の環境	1	2	3	4	5
2	学校の雰囲気	1	2	3	4	5
3	先生の世話のしかた	1	2	3	4	5
4	先生の教えかた	1	2	3	4	5
5	アメリカ人の友達の数	1	2	3	4	5
6	アメリカ人の 友達との交遊の深さ	1	2	3	4	5
7	アメリカ人の友達との会話の量	1	2	3	4	5
8	友達との行動範囲の広さ	1	2	3	4	5
9	クラブ活動	1	2	3	4	5
10	授業内容	1	2	3	4	5
11	あなたの学校での成績	1	2	3	4	5
12	あなたの授業への参加の度合い	1	2	3	4	5
13	あなたの英語能力の進歩	1	2	3	4	5
14	ホストファミリーとの親密度	1	2	3	4	5
15	ホストファミリーの世話のしかた	1	2	3	4	5
16	ホストファミリー での食事	1	2	3	4	5
17	ホストファミリーとの会話量	1	2	3	4	5
18	ホストファミリー での部屋や設備	1	2	3	4	5
19	一般のアメリカ人のあなたに対する態度	1	2	3	4	5
20	あなたの人間的成長	1	2	3	4	5

Research Forum

Evaluating Learner Self-Assessment

Colin Painter
Prefectural University of Kumamoto

This exploratory study examines Pearson product-moment correlations between learner and teacher-assessment in a CAI (Computer Assisted Instruction)-based communicative English course for Japanese university students. It also explores the validation of the program-specific tests used for self-assessment through correlation of the students' self-assessed test scores with their TOEIC scores. Although the self-assessment scores did not correlate significantly with all parts of the TOEIC, significant correlations of self-assessment were observed with teacher assessment, suggesting the reliability of the self-assessment procedure.

本論では、日本の大学におけるコンピュータを利用した英語コミュニケーション授業場面を取り上げ、学習者自身による評価と教師による評価とのピアソン相関係数の分析を行った。さらに、学習者の自己評価得点とTOEIC得点の相関分析を行い、この自己評価に用いたテストの妥当性を検証した。その結果、自己評価点は、TOEICの全部のパート別得点に有意な相関関係があるわけではないが、自己評価得点と教師による評価との間には有意な相関が認められた。このことは、自己評価の信頼性の高さを示唆している。

This exploratory study examines the following aspects of learner self-assessment: (1) whether learner and teacher assessment have positive correlations, thus indicating the reliability of the learners' self-scoring; and (2) whether the role-play tests used for assessment have positive correlations with a standardized test. The study also examines whether the number of self-assessment tests increased compared with the number of teacher-assessed tests reported previously (Painter, 1995).

The following review explores the positive results of studies on learner self-assessment and addresses the necessity of establishing the reliability and validity of the program-specific test used for self-assessment activities.

Learner Self-Assessment

Studies on learner self-assessment are relatively few but report generally positive results. From 1967 to 1998 *TESOL Quarterly* published only one article containing "self-assessment" in the title (LeBlanc and Painchaud, 1985). This paper examined students' ability to self-assess levels in French and English as a Second Language using a questionnaire for placement purposes. Pearson product-moment correlations between a proficiency test and two types of self-assessment questionnaires were .80 and .82. Thus, the authors concluded that self-assessment was valuable as a placement instrument.

Since its founding in 1985, *Language Testing* has published seven papers relevant to the area of self-assessment (Bachman & Palmer, 1989; Blanche, 1990; Heilenmann, 1990; Janssen van Dieten, 1989; Oscarson, 1989; Ross, 1998; Shameen, 1998). One of the most recent (Ross, 1998) includes a meta-analysis of the correlations contained in a number of studies made since 1978 (Bachman & Palmer, 1981, 1982; Blanche, 1990; Buck, 1992; Ferguson, 1978; Janssen van Dieten, 1989; LeBlanc and Painchaud, 1985; Milleret, Stansfield & Mann-Kenyon, 1991; Wongsotorn, 1981). These included research across the four language skills within a wide range of second and foreign language contexts. The criterion Ross employed to select these studies for analysis was the presence of "an empirical basis for evaluating the relationship between self-assessment and a second or foreign language criterion variable" (p. 2). Examining the Pearson product-moment correlations between self-assessment and speaking skills, Ross found the average to be .55 ($p < .05$) for the 29 self-assessments of speaking within the ten studies. Looking at the total of 60 self-assessments across the four language skills, Ross found a correlation of .63 ($p < .05$). Thus, Ross concluded that self-assessment typically offers "robust" concurrent validity with criterion variables.

Other researchers have also made a case for self-assessment. Murphey (1994) noted the ability of a test not only to measure but to stimulate learning. He requested that his students make their own tests and test each other. Believing that there is insufficient time to test everyone orally, he sacrificed teacher control and encouraged students to test each other, inside or outside the classroom.

Computer-assisted Instruction (CAI) is also suggested to engender a learning environment which promotes learner autonomy. Peterson (1997) believes that computer-mediated instruction (CMI) promotes learner autonomy in that it provides a less restrictive learning environment than the traditional language classroom. Citing Cooper and Selfe (1990), Peterson feels CMI is compatible with personal learning styles and encourages the learner to take control of the learning process.

Following the positive views of both self-assessment and CAI, this exploratory study argues for the reliability of student self-assessment made using course-specific tests given in a CAI class for communicative English. Correlational evidence is provided showing a positive relationship with teacher assessment and with some sections of a well-known test of English language proficiency.

Test Types and Criterion-Related Validity

Validity issues usually concern two types of test, Criterion Referenced Tests (CRTs) and Norm Referenced Tests (NRTs). Brown (1995) discusses several characteristics which distinguish CRTs from NRTs, and suggests that the most fundamental is the purpose of the test. He notes that CRTs foster learning and are typically used by teachers to encourage students to study, review, or practice the material in a course. On the other hand, the basic purpose of NRTs is to spread students' performances out so that they can be classified for admission or placement (Brown, 1995, p. 13; 1998). CRTs are more likely used to discover how much of a given level of ability or content domain the test-takers have learned, for example, when a teacher gives a test at the end of a unit of language study. The focus of the CRT, then, is on the relationship between the learner/test-taker *and the material*, whereas the focus of the NRT is on comparing the learners' performances with *one another*.

The CRT, which is based on the syllabus of a course, is likely to have beneficial washback effect on the learners, encouraging them to take the syllabus seriously. After the test, teachers can go through the test questions with the learners, making it a teaching tool. However, NRT test-takers may never learn their mistakes since the NRT paper is less likely to be returned to test-takers. In fact, there may be no direct connection between the multiple-choice questions in the NRT and the syllabus of the course. An important question, then, is whether different CRTs are valid measures of the learners' language skills in general.

Among the different types of validity, *criterion-related validity* is particularly important since it indicates the extent to which scores on one test will estimate or predict performance on other tests measuring the same ability. The primary way of establishing criterion-related validity is by correlating the test in question with another test which is well established and measures the same ability. Although a major issue in test design is the extent to which syllabus-based CRTs can be used as valid indicators of learners' proficiency, Brown (1988, 1995) notes that it is often not possible to use an NRT to validate a CRT since they measure different things, the CRT testing mastery of specific course content and the NRT being a more global measure of language proficiency.

Complicating the validation process of specific CRTs is the lack of a CRT which is well established and is thus appropriately representative of the ability criterion. Bachman (1990) points out that there is a strong need to develop valid criterion-referenced measures of communicative language ability. He feels there is a need for a "common yardstick" (p. 334) and that CRTs would fulfil this need. A recent paper by Nakamura (1995) laments the absence of a relevant CRT which could be used for establishing *concurrent validity* (p. 129), that is, the extent to which results on two tests administered at the same time correlate significantly with each other. He used students' grades in conversation classes and compared them with teacher estimates of their speaking ability to investigate concurrent validity.

Thus, although varied learning situations and their accompanying syllabuses cause difficulties in defining a common level of ability, making the "common yardstick" elusive, both NRTs and CRTs have an important role in program evaluation (Lynch, 1992) and in measuring learning. Mindful of the difficulty of using an NRT to validate CRTs, this exploratory research nonetheless uses an well-known NRT to test the validity of the type of CRT assessment test used in this study.

Validity of the TOEIC

The Test of English for International Communication (TOEIC), developed by The Educational Testing Service (ETS), is an example of an NRT used in language education. Although it does not directly test oral skill, the TOEIC is a well-established language test. MacGregor (1997) suggests that both the TOEIC and the TOEFL are regarded as valid instruments because ETS regularly publishes reliability and validity reports on their use. She cites Wilson (1993) on the link between TOEIC listening scores and the scores on the Language Proficiency Interview (LPI), a direct assessment of oral language proficiency developed by the Foreign Service Institute of the US government. The correlation between the LPI and the TOEIC listening was a consistently high .83, "suggesting that both tests are, as they claim, effective measures of the ability to understand and use spoken English" (p. 32). MacGregor also cites Woodford (1992) who reports that, "in 1989 and 1990, test reliability for TOEIC using the KR-20 formula was .96" (p. 35).

In this report, correlational analysis of learner self-assessment is conducted, using the TOEIC to assess the criterion-related validity of the self-assessment process.

The Study

This exploratory study investigates learner self-assessment during three years of a university CAI oral communication program, 1995-1997. A previous report (Painter, 1995) described how the program aimed at the development of oral communication using computers and how paired learners requested testing through role play after they had completed a unit of functionally-based language activity. The role-play test scores were analyzed for both test-retest reliability and intra-rater reliability (Painter, 1997b) and in both cases the Pearson product-moment correlation coefficient was .88 (p <.05), indicating a significant test-retest correlation (see Painter, 1997b for details). Moreover, test validity was indicated since (1) the ability domain was based on the course outline, and (2) the test scores, as well as the number of tests requested by the students, correlated significantly with cloze test scores (Painter, 1997b). However, it was suggested that further correlation studies of the role-play tests would provide more convincing evidence of criterion-related validity. The participants of the study provided this opportunity when they subsequently took part in the TOEIC, allowing for comparison of the role-play test scores with their TOEIC scores.

Research Focus

Three areas regarding learner self-assessment are explored in this limited report:

(1) Investigation of how self-scored testing affects the pace of learning, as reflected in the number of tests taken during the years of self-assessment compared with the number taken during the period of teacher-assessment.

(2) Investigation of the reliability of the course-specific role-play tests by examining the relationship between learner and teacher scoring.

(3) Investigation of the criterion-related validity of the role-play tests by correlating learner self-assessment scores with a widely used reliable and valid test, the TOEIC.

Method

Participants

Learners at the Prefectural University of Kumamoto, Faculty of Administration are of mixed gender (M:F; 46:54). Classes are ninety minutes in length and the CAI Oral English class is offered once weekly for first-year learners and once biweekly for second-year learners. A total of 151 students participated in this study, and five of the six groups took the TOEIC test, as shown in Table 1.

Description of the Program, Testing, and Test Scoring

The CAI Program

First-year learners begin the CAI program using a situational/functional English software program titled *Nova City, Beginner* (Milward, 1993), containing five units and tests. The units included such topics as "At the Airport," "Checking into a Hotel," and so forth. The second-year learners used the next course in the series, *Nova City, Intermediate*, containing 20 units and tests.

Scoring of the Assessment Tests

The twenty-five performance tests used in the CAI program were CRTs in the form of role-plays derived from the material studied in class (see Painter, 1996, for a full description of the test development process). Pairs of students were requested to perform a role-play based on the material they had just studied. In 1995, the first year of the program, all tests were administered and scored by the teacher. The scoring procedure used during teacher assessment went as follows:

1. Communication was meaningful and grammatically correct:
   2 points for each section
2. Communication was meaningful but contained grammatical errors:
   1 point for each section
3. Communication was meaningless:
   0 points for each section

Table 1: Participants in the Study

Year	Students' year	Number of classes	Learners completing 2 semesters of CAI	Learners taking TOEIC (N= 151)
1995	1st	26	48	22
	2nd	13	48	none*
1996	1st	26	49	29
	2nd	15	43	17
1997	1st	27	47	45
	2nd	16	50	38

*The 1995 second-year learners did not take the TOEIC

Here a "section" refers to a section of dialogue, such as an initiating remark, question, response, or closure. This scoring method attempted to reduce the items the assessor needed to keep track of during the test (Underhill, 1987).

A subsequent study (Painter, 1997b) indicated that learners sometimes had to compete for the chance to test, possibly dampening the positive effects of autonomy and slowing down the assessment process. To learn more about the relationship between performance opportunities and proficiency it was felt necessary to provide unrestrained opportunity for testing. It was thus suggested (Painter, 1997b) that further research should include self-testing and self-grading by learners. This would enable learners to move through the program at their own pace, without any impediment caused by the teacher-administered testing process.

Learner Self-Assessment

Since 1996, learners have graded themselves upon finishing their role-play test at the end of a unit. Since learners were both participants as well as assessors of the test, it was impossible to score sections of the test without interrupting the testing process. Therefore scoring took place after each test. Following the teacher scoring guidelines above, the learners were required to estimate an accuracy level for "Meaningful Communication," then estimate "Grammatical Accuracy." These terms were carefully explained in a guide and exemplified by the teacher at the beginning of the course. The learners were informed that 20% of their final grade would come from the self-assessed test scores.

A one-page English-language *Procedure Guide* was issued to the learners from the first semester in 1995. A revised five-page English-language guide was issued in 1996, and in 1997 the *Procedure Guide* was issued bilingually (Painter, 1997a).

Correlational Analysis

For the purpose of comparison between learner and teacher-assessment, simultaneous scoring began in 1996. Twenty-three categories were used for analysis, as shown in Figure 1. Some categories, such as "grade" and its components such as "attendance," are self-correlated. However, in the interest of comprehensive investigation, all categories were recorded for comparison. Spreadsheets with Pearson's product-moment correlation matrixes were produced representing the data from each of the learner groups. Only a small portion of this data is generated for the present report.

The learners' TOEIC test results were used for the purpose of comparing self-assessment with a validated test. Data was recorded over the six semesters covered by the study, 1995-1997. Two groups of first-

Figure 1: Correlation Categories

1. Learner self-assessed performance (1 time only, 7/1996)
2. Teacher scored performance (1 time only, 7/1996)
3. TOEIC listening score
4. TOEIC reading score
5. TOEIC overall score
6. Cloze score, first semester
7. Cloze score, second semester
8. Cloze score, average
9. Learner self-assessed average performance score, first semester
10. Learner self-assessed average performance score, second semester
11. Learner self-assessed average performance score
12. Performance test quantity, first semester
13. Performance test quantity, second semester
14. Performance test quantity, total
15. Homework quantity, first semester
16. Homework quantity, second semester
17. Homework quantity, total
18. Attendance, first semester
19. Attendance, second semester
20. Attendance, average
21. Grade, first semester
22. Grade, second semester
23. Grade, average

year learners were studied in both semesters of 1995. However, the TOEIC was not taken by the 1995 second-year learners, therefore only basic data appears for them. Two groups of first and second-year learners were studied in both semesters of 1996. Also, two groups of first and second-year learners were studied in both semesters of 1997. The data for TOEIC-takers from identical learner-year groups is combined for the purpose of the correlation study. Pearson product-moment correlation matrixes were made for all learner groups. The data contained in the tables below is derived from the matrixes, and a descriptive statistics table appears in the Appendix. Space limitation prevents the display of the matrixes themselves.

Results

Test Quantity and Self-Assessment

During 1995, the period of teacher-assessment, the first-year learners took an average of nine assessment tests, these scored by the teacher

(Table 2). In 1996, with self-assessment, there were 12 tests per first-year learner, an increase of 33%, and in 1997, these learners took 13 tests. Interestingly, the average score of tests remained the same, at about 79%, regardless of whether assessment was made by the teacher or the learners. Second-year learners receiving teacher assessment took only four tests, but when conducting self-assessment in 1996, they took an average of six tests, with an average score of 75%, an increase in output of 50%. The average scores of the 1997 second-year learners were almost the same at 77%, while test quantity was the same, at six tests during the year. Thus, both first- and second-year learners took more tests when self-assessing, and the self-assessment procedure did not appear to result in inflated scoring.

Table 2: Influence of Self-Assessment on Test Quantity & Average Score

Year	Year	Average Test Score**	Number of Tests Taken**
1995*	1st	79	9
1996	1st	79	12
1997	1st	80	13
1995	2nd	74	4
1996	2nd	75	6
1997	2nd	77	6

* Only teacher-assessment was used in 1995

** Values for test scores and number of tests taken have been rounded

Teacher and Learner Assessment Compared

In the first semester of 1996, 68 tests were scored simultaneously, both by learner self-assessment and by the teacher. To compare the reliability, a one-time correlational analysis of self-assessment and teacher-assessment using the tests given in July, 1996 was performed, and the results are shown in Table 3. First-year learner self-assessment and teacher-assessment correlated significantly at .53 ($p < .05$). The correlation of $r = .66$ ($p < .05$) for the second-year assessments was also significant.

Correlational Analysis of Learner Assessment Scores with the TOEIC

Table 4 shows first-year and second-year learners' scores correlated with the TOEIC for 1996 and 1997, first-semester and second-semester tests, and the two sets of scores for each year combined and recorrelated.

Table 3: One-Time Correlation of
Learner Self-Assessment and Teacher-Assessment

Year	Year of Study	Number of Students	Correlation
1996	1st	29	[illegible]
	2nd	17	.66*

* Significant ($p < .05$)

In the first semester of 1996, the first-year learners' self-assessment indicated a weak non-significant correlation with TOEIC Overall, as shown in Table 4 below. However, the second-year learners' scores had significant correlations with TOEIC Listening, Reading and Overall Total, at $r = .46$ ($p < .05$), $r = .42$ ($p < .05$) and $r = .54$ ($p < .05$) respectively.

The second-year 1997 learners' TOEIC scores dated from 18 months prior to their participation in the CAI program, and there was no significant correlation between those scores and the scores obtained in the program (Table 4). However, for the first semester of 1997, the first-year learners' self-assessment average correlated significantly with both TOEIC Listening, at $r = .35$, and TOEIC Overall Total at $r = .29$.

Only eight significant corrrelations out of 36 were observed between the TOEIC and the self-assessment scores of the learners, with three of the eight coming from the larger number of tests represented in the combined first and second semester scores. Therefore, the validity of learner self-assessment receives only slight support from correlation with the learners' TOEIC scores.

Table 4: Correlation of Self-Assessed Average Performance Scores with TOEIC

Year	1996						1997					
Learner year of study	First			Second			First			Second		
Semester of self-assessment	1	2	1+2	1	2	1+2	1	2	1+2	1	2	1+2
N	29	29	29	17	17	17	45	45	45	38	38	38
TOEIC listening	.22	.18	.24	.30	.46*	.41*	.35*	.24	.30*	-.06	.05	.01
TOEIC reading	.13	.28	.25	.29	.42*	.38	.17	.08	.13	-.02	.19	.09
TOEIC total	.18	.26	.27	.36	.54*	.48*	.29*	.18	.24	-.06	.13	.04

*Significant ($p < .05$)

Discussion

In the CAI program, completing a unit of study was a pre-condition for taking a role-play assessment test. Consequently, the number of tests taken implies the pace of study. With sizeable groups of learners, having the teacher assess every learner pair's role-play is impractical and is believed to slow down the learners' progress (Painter, 1997b). In this program, the transition to self-assessment resulted in an increased pace of learning without an accompanying inflation of grades through the self-scoring procedure. The increase of between 33% and 50% in the number of tests taken, with stability of scoring maintained, observed under self-assessment suggests that self-assessment has a positive influence on the pace of learning.

However, the increased number of tests taken without inflated self-grading, in itself, is not sufficient to establish the reliability of the self-assessment procedure. It is also desirable that learner self-assessment be significantly correlated with teacher-assessment. In this study, first-year and second-year learner self-assessment scores on one test correlated significantly with teacher-assessment, suggesting reliability in self-assessment. Clearly, however, wider correlational studies are necessary.

Concerning validity, self-assessment was examined for correlation with the TOEIC, a validated NRT. As noted, the purposes of NRTs such as the TOEIC, and CRTs, which are program-specific tests measuring learner mastery of what has been taught, are quite different and one should not necessarily expect significant correlations. In this study, only a few significant correlations were observed. Further research is also necessary in this area.

Conclusions

The results of this exploratory study suggest that self-assessment enhances the output of performance while retaining stability of scoring. Reliability of the self-assessment process was suggested by the significant correlation between learner and teacher scoring procedures on a single test. Only limited confidence, however, is suggested concerning the criterion-related validity of the self-assessment test due to the small number of significant correlations between parts of the TOEIC and the self-assessed role-play tests.

Further research should consider the need for larger groups, perhaps assembled by combining results from several classes of learners being taught by similarly interested teachers. A training period would be necessary in which learners are first tested on their grasp of the criteria for self-assessment, followed by a period to harmonize their self-assessment ratings. In this way, reliable results could be produced from subsequent correlation studies. Teacher-researchers are encouraged to try out self-assessment in their teaching situations.

The learners in this study were certainly enthusiastic about the opportunity to assess themselves and the washback effect was evidenced by the 33%-50% increased output noted. Tying self-assessed scores to a modest percentage of the grade, such as the 20% in this study, convinces learners that they are being taken seriously.

Acknowledgements

This is a version of a paper presented at the Japan Association of College English Teachers (JACET), 36th Annual Convention Program, Waseda University, Tokyo. The author is grateful for advice given at the beginning of the program, particularly by Dr. Thomas Robb, Chair, English Department, Kyoto Sangyo University and Dr. John Shillaw, Tsukuba University. Thanks are due to the two anonymous JALT Journal reviewers for their valuable suggestions, as well as to the students who participated in the study. Gratitude is expressed toward colleagues for their support.

Colin Painter is an Associate Professor at the Prefectural University of Kumamoto. He has taught at universities in Asia for the last 16 years. His interests include language acquisition, curriculum development, and computer-assisted language learning.

References

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A. (1981). The construct validity of the FSI Oral Proficiency Interview. *Language Learning*, 31, 67-86.

Bachman, L.F. & Palmer, A. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16 (4), 449-65.

Bachman, L.F. & Palmer, A. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6 (1) 14-29.

Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing*, 7 (2), 202-229.

Blanche, P & Merino, B. (1989). Self-assessment of foreign language skills: Implications for teachers and researchers. *Language Learning*, 39, 323-340.

Brown, J.D. (1988). *Understanding research in second language learning*, Cambridge: Cambridge University Press.

Brown, J.D. (1995). Differences between norm-referenced and criterion-referenced tests. In J.D. Brown & S.O. Yamashita (Eds.). *Language Testing in Japan* (pp. 12-19). Tokyo: The Japan Association of Language Teaching.

Buck, G. (1992). Listening comprehension: Construct validity and trait-characteristics. *Language Learning*, 42, 313-57.

Cooper, M.M. & Selfe, C.L. (1990). Computer conferencing and learning: Authority, resistance and internally persuasive discourse. *College English*, 52 (8), 847-869.

ETS (Educational Testing Service). (1992). *Guide to SPEAK*. Princeton, NJ: Educational Testing Service.

Ferguson, N. (1978). Self-assessment of listening comprehension. *International Review of Applied Linguistics*, 16, 146-156.

Heilenmann, L.K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7 (2), 174-201.

Janssen van Dieten, A. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6 (1), 30-46.

LeBlanc, R. & Painchaud, G. (1985) Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19 (4), 673-687.

Lynch, B. (1992). Evaluating a program inside and out. In J.C. Alderson & A. Beretta (Eds.). *Evaluating second language education* (pp. 61-99). Cambridge: Cambridge University Press.

MacGregor, L. (1997). The *Eiken* test: An investigation. *JALT Journal*, 19 (1), 24-42.

Milleret, M., Stansfield, C. & Mann-Kenyon, D. (1991). The validity of the Portuguese speaking test for use in a summer study abroad program. *Hispania*, 74, 778-787.

Milward, M. (1993). *Nova City*. (CD-ROM) Tokyo: Nova Information Systems.

Murphey, T. (1994). Tests: learning through negotiated interaction. *TESOL Journal*, 4 (2), 12-16.

Nakamura, Y. (1995). Making speaking tests valid: Practical considerations in a classroom setting. In J. D. Brown & S. O. Yamashita (Eds.). *Language Testing in Japan* (pp. 126-133). Tokyo: The Japan Association of Language Teaching.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6 (1),1-13.

Painter, C. (1995). Developing oral communication using computers: Computer assisted language learning. *Administration*, 2 (3), 109-150.

Painter, C. (1996). *Performance Tests*. Kumamoto: Prefectural University of Kumamoto, Foreign Language Education Center.

Painter, C. (1997a). *Procedure Guide For Using Software* (Bilingual) Mimeograph. Kumamoto: Prefectural University of Kumamoto, Foreign Language Education Center.

Painter, C. (1997b). Continuous assessment facilitated by CAI. In S. Cornwell, P. Rule & T. Sugino (Eds.). *On JALT 96, Crossing Borders* (pp. 119-125). Tokyo: The Japan Association for Language Teaching.

Peterson, M. (1997). Language teaching and networking. *System*, 25 (1), 29-37.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15 (1), 1-20.

Shameen, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing,* 15 (1), 86-108.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques.* Cambridge: Cambridge University Press.

Wilson, K. (1993). Relating TOEIC scores to oral proficiency interview ratings. *TOEIC Research Summaries 1.* Princeton, NJ: Educational Testing Service.

Wongsotorn, A. (1981). Self-assessment in English skills by undergraduate and graduate students in Thai universities. In J. Read (Ed.) *Directions in language testing* (pp. 240-260) Singapore: Singapore University Press.

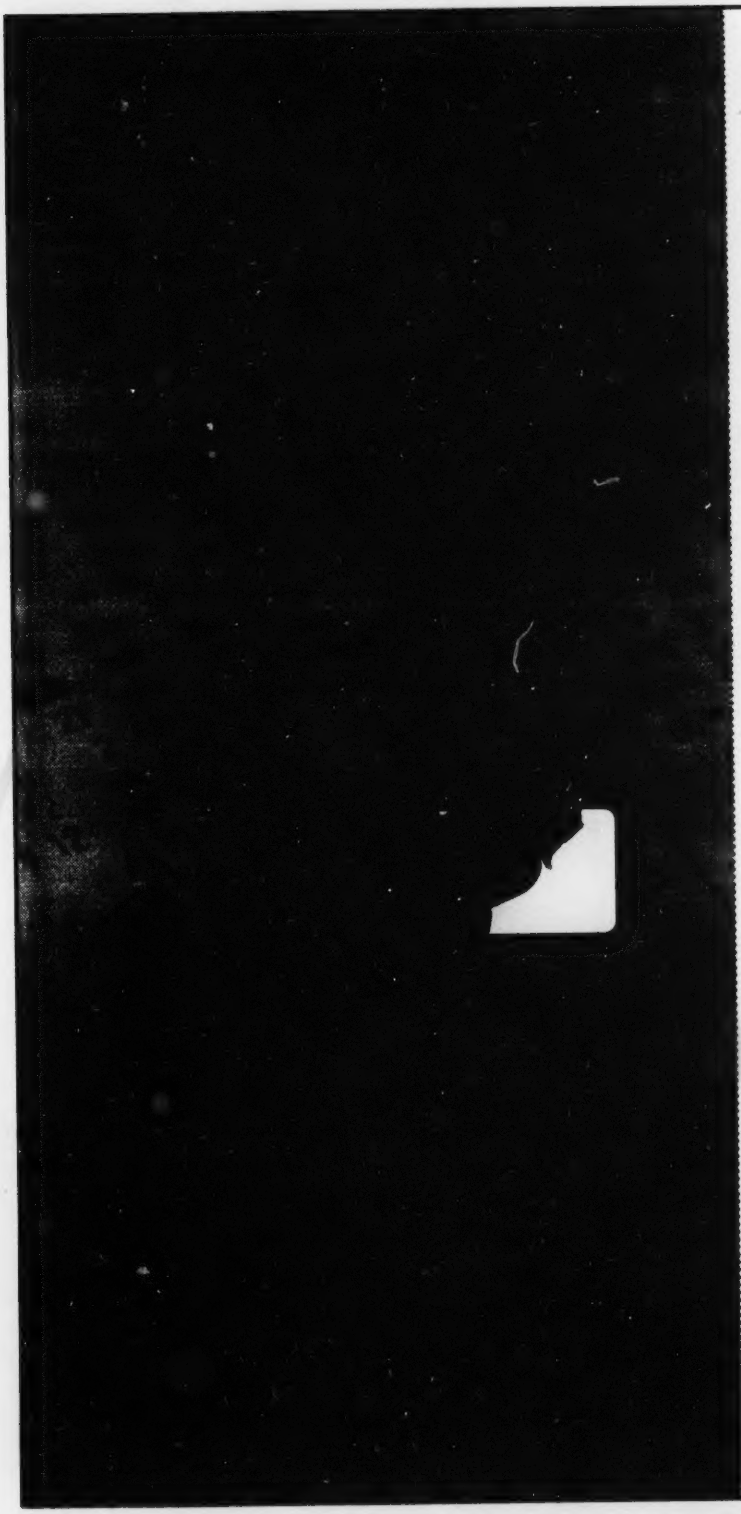
Woodford, P. (1992). A historical overview of TOEIC and its mission. *The 35th TOEIC seminar* (pp. 10-15). Tokyo: The Institute for International Business Communication.

(Received October 5, 1997; revised December 21, 1998)

Appendix: Descriptive Statistics Table

		1 L Perf 1 time	2 T Perf 1 time	*3 TOEIC Listen	4 TOEIC Read	5 TOEIC Total	6 Cloze sm 1	7 Cloze sm 2	8 Cloze Ave	9 L Perf sm 1	10 L Perf sm 2	11 L Perf Ave	12 Perf 1 tests	13 Perf 2 tests	14 Perf Total
1995-6 First Year	n			22	22	22	22	22	22	22	22	22	22	22	22
	Administered	N/A	N/A	10/96	10/96	10/96	7/95	12/95		~7/95	~1/96		sm 1	sm 2	
	SD			39.0	44.9	64.5	6.9	8.4	6.6	8.4	9.4	7.1	0.7	0.6	1.1
	Mean			194	135	329	85	73	79	80	78	79	4	4	9
1996-7 1st Year	n	29	29	29	29	29	29	29	29	29	29	29	29	29	29
	Administered	7/96	7/96	5/96	5/96	5/96	7/96	12/96		~7/96	~1/97		sm 1	sm2	
	SD	16.0	17.0	30.2	48.3	72.4	9.0	7.9	7.8	8.7	6.9	6.8	1.0	1.5	2.1
	Mean	71	77	197	155	352	80	72	78	76	82	79	7	6	12
1996-7 2nd Year	n	17	17	17	17	17	17	17	17	17	17	17	17	17	17
	Administered	7/96	7/96	10/96	10/96	10/96	7/96	12/96		~7/96	~1/97		sm 1	sm2	
	SD	15.2	16.1	47.7	45.7	76.7	7.2	11.4	8.4	11.1	13.1	11.5	0.7	1.1	1.5
	Mean	69	83	217	164	382	53	54	53	75	75	75	3	3	6
1997 1st Year	n			45	45	45	45	45	45	45	45	45	45	45	45
	Administered	N/A	N/A	5/97	5/97	5/97	7/97	12/97		~7/97	~1/98		sm 1	sm2	
	SD			46.5	49.7	85.9	7.1	12.2	8.4	8.8	10.6	9.3	1.0	1.5	1.9
	Mean			221	150	371	87	75	81	78	81	80	6	7	13
1997-8 2nd Year	n			38	38	38	38	38	38	38	38	38	38	38	38
	Administered	N/A	N/A	5/96	5/96	5/96	7/97	12/97		~7/97	~1/98		sm 1	sm2	
	SD			43.2	34.9	64.8	12.1	11.8	10.4	7.6	5.7	5.8	0.8	0.9	1.6
	Mean			209	172	381	62	59	60	76	78	77	1.9	3.9	5.6

sm: Semester
L: Learner
Perf: Performance
Tests: Test Quantity
*TOEIC administration may pre- or post-date year of study



For the first time retail investors can access a unique investment opportunity previously only available to institutional investors.

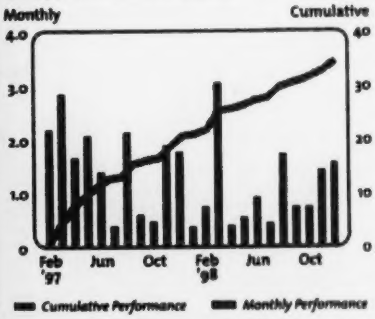
The Paradigm Fund – a superior retail investment product from Banner Japan – accesses expert management to realise the return potential of a formerly exclusive investment sector.

The strategy underpinning the Fund has demonstrated solid, positive returns since inception through varying investment conditions.

Investing only in AAA-rated US Mortgage Backed Securities, the Fund seeks to generate high returns through sophisticated management of this sector's unique risk and return profile.

Utilising the cutting-edge analytics of a pre-eminent US investment house, the Fund aims to deliver a dynamic, risk-controlled investment strategy and a tax effective investment.

Strategy Performance (% increase)



Note - past performance is no guarantee of future returns.

For a brochure, or more information, contact:
Banner Overseas Financial Services
Tel (03) 5724 5100 Fax (03) 5724 5300
Email banner@gol.com
www.paradigmfund.com

BANNER

BEST COPY AVAILABLE

Perspectives

Raising the Quality of Discourse Using Local Area Networks in Returnee Classes

John Herbert
Ritsumeikan University

A well-designed computer local area network (LAN) can act as a valuable tool in the second language classroom. This paper looks at the ways in which one such LAN has been put to use in a returnee class in a Japanese university. The paper asserts that the quality of discourse is raised in the computer-assisted classroom discussion for several reasons. These reasons include: (a) Students can work at their own pace; (b) many students can take part in a synchronous discussion; and (c) students are more willing to self-disclose in a computer-assisted discussion than might be expected in a traditional oral setting. The results of a series of LAN discussions conducted in a returnee class, along with feedback from students, are used to provide analysis of this technique.

洗練されたコンピュータ・ローカル・エリア・ネットワーク(以後LAN)の利用は、外国語教育にとって有効な手段となり得る。本研究では、日本の大学の帰国子女クラスにおけるＬＡＮ利用について報告する。コンピュータを利用することにより、学習者の談話に質的な向上が認められることが本研究の論旨であるが、このような質的向上の理由には、ａ）学習者が各自のペースで学習できること、ｂ）多人数の学習者が同時にディスカッションに参加できること、c) 従来型の授業形態と比較して、コンピュータを利用した授業では、学習者はより自己表現を行う、の三つの理由が挙げられる。ＬＡＮ上で実際に行われたディスカッションや学習者からのフィードバックをもとに、この教授法の有効性を検証する。

The teaching of English as a second language has been affected by the computer industry and it is common for English programs in many educational institutions to make use of the computer as a resource for second language learning. Before the 1990s most of the software involved fairly simple reading, grammar or word processing programs but since the turn of the decade, computer networks have been utilized in the classroom. As opposed to the international networks that make use of the Internet to allow people to interact through electronic mail and MOOs (Multiple-user-domain Object Oriented) (see Davies, Shield, & Weininger, 1998), local area networks (LANs) can be confined to one classroom and do not require access to the World Wide Web. Utilizing a well-designed LAN enables large numbers of students to take part concurrently in a real-time discussion in a computer classroom setting without the practical complications associated with accessing the Internet.

Computer-Assisted Classroom Discussions

Computer-assisted classroom discussions (CACDs) have several well-documented advantages over traditional oral classroom discussions. Ortega (1997) identifies the following positive results emerging from research on CACDs: (a) an equalizing effect on learner participation in discussions (Beauvois, 1992; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996; Warschauer, 1996); (b) increased learner productivity, with implications for second-language (L2) acquisition considering that practice in production of the L2 promotes transformation from L2 learning to L2 acquisition (Stevick,1986, as cited in Larsen-Freeman & Long, 1994); and (c) the tendency for the quality of language produced in CACD to be more complex than that produced in face-to-face discussions (Warschauer, 1996).

Following this last finding, this exploratory report will discuss discourse quality and participation in a CACD forum. Since quality of discourse is very difficult to define, this paper will not address the topic in terms of a quantitative study of linguistic accuracy, but rather will look at the nature of the English output produced by students in the electronic format through quotations and interpretation. It will be argued that, in holistic terms, the quality of discourse produced in CACD is raised for the following reasons: (a) students work at their own pace; (b) they can swap opinions in a discussion forum in large numbers; and, (c) as Ma (1996) has noted, they are more willing to self-disclose in the computer-mediated discussion format than they are in face-to-face discussions.

Working at Their Own Pace

The use of LANs for computer-mediated discussion allows students to work at their own pace. In an oral situation a student is under pressure to answer questions within a certain time, whereas in CACD a student has time to formulate ideas and can read the opinions of others before composing and sending a message. This lack of time pressure acts in several positive ways to produce a higher quality of discourse.

First, those students who may be reticent in oral discussions due to time-pressure anxiety tend to play a greater role in class discussions. Equalizing participation produces a wider based discussion that allows students to access the views of all their peers, not just the more dominant students.

Second, without the necessity to reply immediately, students in a CACD can spend time formulating their ideas before communicating them to the class. Self-monitoring of their written messages, stressed as a key component in thinking and communicating (Slatin, 1991, cited in Markley, 1992), can also take place, allowing students to make changes to their work in the editing window of the computer screen before sending their comments to their peers.

Facilitating Interaction

In a traditional oral discussion class, the teacher is faced with a logistical dilemma. Whole-class discussion is often time-inefficient since students must listen to the opinions of the student who is speaking and wait for their opportunity to give their views. The solution is to divide the class up into small groups. (For a comparison of small-group oral discussions with networked computer discussions see Freiermuth, 1998). However, group work has several negative effects on the quality of the discussion.

First, the wide-based aspect of the discussion is lost since the audience is limited to only a few students. In CACDs, however, students can consider a wide range of views and find a strand of discussion or sub-issue that interests them. They can then develop this topic with others who have the same interests, forming a small group based on interest.

Second, a teacher may have difficulty in monitoring all students' output in a small group discussion, whereas in CACD the teacher is in contact with all students through the computer screen. This allows the teacher to guide the discussion in order to help the students delve deeper into the issues.

Third, since all comments made by students appear on the upper half of the computer screen, students have the option of using the scroll bar to review the messages sent during the class. This is an advantage over the small-group format in that students may refer to arguments or opinions given previously. This is only possible in the oral format by interrupting the flow of discussion and checking on opinions or comments made several minutes earlier.

Greater Willingness to Self-Disclose

Based on a study of synchronous "relay" sessions conducted between US students and East Asian students (60% of whom were studying in US universities), Ma (1996) claims that both East Asians and North Americans have a tendency to show greater self-disclosure in CACDs than in face-to-face oral discussions. Ma (1996) uses Berger and Calabrese's (1975) uncertainty reduction theory to describe self-disclosure as being "willing to proffer information about themselves without specifically being asked for it" (Ma, 1996, p. 178), including personal opinions or feelings. Ma's findings show that whereas both sets of students perceived themselves as showing greater self-disclosure, almost half of the US students did not feel that the East Asians self-disclosed more in the computer-mediated mode than in face-to-face conversations.

Research Focus

In this exploratory investigation, self-disclosure is defined as willingness to disclose information about oneself and to give personal opinions that further reveal information about oneself. The research focus of this study was to determine whether Japanese university "returnee" students would participate and self-disclose using CACD. This paper does not present a quantitative analysis of data, but rather shows extracts which suggest the degree of self-disclosure and discourse quality, and presents selected results of a questionnaire on participation in the online discussions.

Method

Participants

The participants were thirty-five students, aged 18-20, taking a Reading and Writing class at a Japanese university. Eighteen were female and 17 were male, with TOEFL scores ranging from 480 to 640. All had spent time in educational systems outside of Japan, with an average length abroad of three years. Such students are usually referred to as "returnees" in Japan.

Materials

The Interchange application of the *Daedalus Integrated Writing Environment* (DIWE) (1994) was used in the returnee class. DIWE runs on Macintoshes or PC-compatibles, and the software enables the linking of computers to form a network. The Interchange application can be found within this software package and is easily accessed by students from the "message" menu once they have logged onto DIWE. After completing this step, students are presented with a screen that is split horizontally into two windows. In the lower window, students type their contributions to the discussion and click on the "send" button. All messages appear in the top window in the order they were sent, with the sender's name above each message. Students can view the full contents of the top window at their own pace using the scroll bar.

For the first CACD presented here, the students read an article on bullying from a website newspaper (*The Times*, 1997) prior to the session. The second session used teacher-generated material dealing with prejudice and discrimination. At the end of the course, students were given a questionnaire to complete relating to the CACD classes. Nineteen responses to the questionnaire were returned.

Procedure

The participants spent the second semester of the Reading and Writing course discussing various issues using the Interchange function of DIWE. Before each class the students were assigned the material to read. This material provided the basis for CACD in the following class. Students were encouraged to give their opinions on the issues raised and were told that participation was expected from all. Students had between fifty minutes and one hour to contribute to the discussion. Discussion questions based on the readings were assigned at the beginning of CACD and were worded in such a way as to encourage self-disclosure, but also to allow students to avoid self-disclosing if they felt inhibited by the subject matter. These questions appeared at the top of the students' computer screens. Students were told that their CACD participation would make up part of their grade for the semester. Extracts from two of the classes are presented and discussed below.

Results and Discussion

The following are short extracts taken from the Interchange CACD conducted on two different class days during the semester. For reasons of anonymity, students' names have been abbreviated. The extracts have not been corrected for mistakes.

The First Discussion

In Week Three of the semester, the students were assigned an article on bullying in British schools (*The Times*, 1997) in which two adults, one of whom had been a bully and the other the victim of bullying, shared their experiences of school life. The teacher posed the following question: "Tell us about your experiences and stories of bullying. This may be a case that involved you or it may have been a case that you saw or heard about. Why do you think the person in that case was bullied?" This appeared at the top of the students' computer screens. Below are two messages from the discussion.

K.S.: When I was 2nd grade, my class was 31 student. The boys were 21 and the girls were only 10 student. In my class, one girl was bullied. She was always alone from one day. I really didn't know why she was bullied, but I didn't play with her. The other 9 girls including myself were always together, and we ignored her like she was not in there. At that time, I couldn't feel and think how she was got a shock and sad. I believed that she wasn't nice to me and she had been mean so she was bullied. At that time, we were too young to think and care all of things. I think difference was a biggest problem for us.

R.Y.: I bullied the girl in my class, because everyone in my class did the same thing, so I didn't feel sorry about her at that time. But when I think back about that time, I think I was doing really stupid thing. Fortunately, the girl who was bullied was strong, so she came to school everyday and acted she was fine, but if she was mentally weak, it was possible that she killed herself because we bullied her. People need to be mature enough to understand how bullied feel.

The discussion involved more than thirty students and the two extracts give a flavor of the form that the discussion took. The students were able to formulate what they wanted to write before sending their comments to their peers. One student wrote on her questionnaire, "When you speak, especially [in a] foreign country, your thinking is sometimes not pretty much composed. On the other hand, when you use CACD you can check out what you are going to say, so it is [a] very good device for discussion."

The Second Discussion

In Week Six of the semester, students were assigned teacher-generated material dealing with prejudice and discrimination. Due to the large volume of written material produced in previous CACDs, students were given a choice of three separate CACD forums. The most popular choice dealt with the topic of gay rights. The discussion question was, "Should gays be allowed to be officially married and enjoy the rights that heterosexual couples receive?" The question itself did not call for self-disclosure as had been the case in the CACD on bullying, although the opinions of the students were sought. The first two messages appeared early in the discussion and are good examples of opinion-swapping at a localized level within the whole-class environment. The last message appeared towards the end of the discussion.

J. K. to M. S.: do you really agree with gay marriage? don't you have any prejudice? i do have prejudice to all homosexual. it's not the original way, isn't it?

M. S. to J. K.: I don't have prejudice to any homosexuals. I have so me gay friends and they are nothing different. Why do you have prejudice to them?

M.Y.: I think we are free to love the others, so it has to be O.K. that gays get marriged (sic). I had friends who were gays when I was in the US. it was my first time to meet or get friends with gays. When I found out that they were gays I was shocked and scared, because we were friends and living together in the girls dorm. She liked one girl who was also my friend and she was a gay also and they had been together about a year or so. It really surprised me, but she talked with me about all this. I realized that it seemed different way of love, but it is same and we do not have right to stop them loving.

Universal Participation and Self-Disclosure

Every student took part in the discussion on bullying, and with only one exception, all made at least two messages. One student observed, "the people who usually didn't participate in class discussions were more active in CACD class. CACD allowed us to think and conclude our thoughts without any time limits, so it gave everybody an equal chance to participate."

CACDs allowed a flow of opinions and expression of a variety of views. One student commented, "[I got] the opportunities to know opinions of other students which I otherwise would never have known, by virtue of CACD's effect of enabling people to have a time to calm down and to take into considerations as much variety of opinions as possible on their display at a time before giving a response." In both discussions, all students participated, with four to five messages being the norm. That breadth of discussion may not have been possible in a small-group oral discussion and would only have been possible in a time-inefficient manner in a full-class oral discussion. It should be noted, however, that time on task is longer in CACD format than in small group discussions. That may be seen as an advantage by some, a disadvantage by others (e.g., Freiermuth, 1998).

When asked to compare self-disclosure in CACD classes with self-disclosure in a spoken classroom discussion, 79% of the respondents agreed that they found it easy to self-disclose in the CACD, with only 10% disagreeing. When students were asked whether they felt that the other students self-disclosed more in CACD than they would have verbally, 74% agreed that their peers showed more self-disclosure in CACD format, and not one student disagreed.

Implications

It is important to state that this paper does not advocate the replacement of oral discussion classes with LAN computer discussion classes. Rather, the computer-mediated discussion format is suggested to be an additional pedagogic resource that will help to enhance an English program.

The discussion classes held in CACD format are suggested to have produced discourse of greater quality than that produced by the same group of students in an oral class, and also to have enabled even the shyest students to participate. However, to achieve this positive result, it was necessary to inform students that they were required to participate and to encourage them to give their opinions and explain their reasons for holding those opinions. When these instructions were given, a wide-ranging flow of opinions ensued. Students who were usually dominant were less so in the CACD, and those who tended to be reticent contributed far more in the electronic domain. It was commonplace for students to personalize the issues they were considering, and self-disclosure took place even when the question that had been posed did not directly require it.

Conclusion

There are many factors that influence the quality of discourse that have not been examined in this exploratory study. The choice of topic will, as Reid (1991) shows, have great bearing on a student's performance. Furthermore this holistic interpretation makes no attempt to provide a quantitative analysis of CACD discussions or to contrast them with the results of small-group oral work. However, having observed the performances of students in both CACD and small group format, this researcher suggests that greater self-disclosure took place in CACD format. Not only were students able to become more aware of the issues being discussed when those issues were personalized, but their willingness to self-disclose also showed an uninhibited spirit, which in turn, allowed a freer flow of opinions among students. This free flow of opinions, coupled with large numbers of students working at their own pace in a concurrent CACD, helped to create a higher quality of discourse. Clearly, future empirical studies of CACDs are necessary to examine both quality and quantity of discourse.

Acknowledgments

I owe special thanks to Professor Phillip Markley, as well as to the JALT Journal*'s anonymous readers, for their helpful critiques of earlier drafts of this paper. I would also like to thank Professor Bernard Susser for help in supplying some of the background literature.*

John Herbert is an English lecturer in the International Relations Faculty of Ritsumeikan University and is currently carrying out research in the use of LANs in the foreign language classroom. He has taught English in Japan for seven years.

References

Beauvois, M. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals*, 25 (5), 455-464.

Berger, C.R. & Calabrese, R.J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1, 99-112.

Daedalus integrated writing environment (Computer software). (1994). Austin, TX: The Daedalus Group, Inc.

Davies, L.B., Shield, L. & Weininger, M.J. (1998). Godzilla can MOO, can you? MOOs for construction, collaboration & community and research. *The Language Teacher*, 22 (2), 16-21.

Freiermuth, M.R. (1998). Using a chat program to promote group equity. *CAELL Journal*, 8 (2), 16-24.

Kelm, O. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. *Foreign Language Annals*, 25 (5), 441-545.

Kern, R. (1995). Restructuring classroom interaction with network computers: Effects on quantity and characteristics of language production. *The Modern Language Journal*, 79, 457-476.

Larsen-Freeman, D. & Long, M.H. (1994). *An introduction to second language acquisition research*. London: Longman.

Ma, R. (1996). Computer-mediated conversations as a new dimension of intercultural communication between East Asian and North American college students. In S.C. Herring (Ed.). *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 173-185). Amsterdam & Philadelphia: John Benjamin.

Markley, P. (1992). Creating independent ESL writers & thinkers: Computer networking for composition. *CAELL Journal*, 3 (2), 6-12.

Ortega, L. (1997). Processes and outcomes in networked classroom interaction: Defining the research agenda for L2 computer-assisted classroom discussion. *Language Learning and Technology*, 1 (1), 82-97.

Reid, J. (1991). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.). *Second language writing: Research insights for the classroom* (pp. 191-210). Cambridge: Cambridge University Press.

Slatin, J. (1991). Forum sessions: Real-time on-line communication and the formation of a discourse community. The computer as instructional medium in an English composition classroom. Unpublished manuscript.

Sullivan, N. & Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System*, 29, 491-501.

The Times (1997). Bullying—The dark shadow over childhood. Aug 25. WWW document: URL <http://www.the-times.co.uk/cgi-bin/BackIssue?1729277>

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal*, 13, 7-25.

Received March 2, 1998; revised May 3, 1998)

The Relationship between Self-Efficacy and Language Learners' Grades

Stephen A. Templin
Meio University

This research explores the hypothesis that students with high self-efficacy: high beliefs in their capabilities to accomplish a task, will achieve higher grades in second language classes than students with low self-efficacy. Seventy-four Japanese high school students were asked to fill out a questionnaire and indicate by a yes or no response which grades they thought they could attain. They also rated their degree of confidence as a percentage for each level. Participants' scores were the total of confidence percentages for "yes" answers. In estimating reliability, Cronbach's alpha for the questionnaire and its subsections was .96, .98, and .91 respectively. A t-test was used to determine if there was any significant difference between low and high self-efficacy students' grades. High self-efficacy students achieved significantly higher grades than low self-efficacy students.

本研究では、「自己効力感が高く、課題の達成能力に強い自信を持つ学生と、そうでない学生とを比較し、第二言語授業においてより高い成績を収める」という仮説を検証する。74名の日本人高校生を対象に、質問紙法を用いて、課題達成に対する自己効力感・達成可能性・自信を10段階評価法・Yes-No評価法、%評価法で評価させた。次に質問への自信度を合計して自己効力感得点として分析を行った。質問紙全体とパート毎の信頼性を示すCronbach's alphaは、それぞれ.96、.98、.91と高い値を示した。自己効力感の高い生徒群と低い生徒群の間で、成績を従属変数にしたティ検定を行ったところ、5%以下の危険率で統計的有意差が確認され、上記の仮説が検証された。

Self-efficacy is belief in how well one can accomplish tasks. Although self-efficacy studies have appeared frequently in psychology (Bandura, 1986; Lee & Bobko, 1994; Locke & Latham, 1990) and management research (Gist & Mitchell, 1992; Gist, Schwoerer & Rosen, 1989; Matsui, Ikeda & Ohnishi, 1989; Matsui & Tsukamoto, 1991), self-efficacy research in second language acquisition (SLA) is rare.

Self-efficacy is important because it influences an individual's performance in two ways. First, a person with high self-efficacy towards a

JALT Journal, Vol. 21, No. 1, May, 1999

task pays more attention, makes a greater effort, is more persistent, and uses a greater variety of strategies to accomplish a task than one with low self-efficacy (Earley & Lituchi, 1991; Lee & Bobko, 1994). High self-efficacy individuals attribute failure to internal causes more than low self-efficacy individuals, who prefer to blame external events (Earley & Lituchi, 1991; Lee & Bobko, 1994). Consequently, when those with high self-efficacy encounter obstacles, setbacks, and failure, they will increase their attention, effort, persistence, and strategies in order to accomplish the task. In contrast, those with low self-efficacy are more likely to give up when faced with similar obstacles.

Second, highly efficacious people actively seek challenging goals and these goals lead to increased performance (Bandura, 1986, p. 391; Griffee, 1997a; Griffee & Templin, 1998). Inefficacious people avoid challenging goals that they fear will lead to negative outcomes. As a result, they do not perform as well.

Other Self-Phenomena

Self-efficacy is not exactly the same as other self-phenomena such as self-concept, self-esteem, confidence, and self-confidence (Ellis, 1990; Griffee, 1997b; Heyde, 1979; Larsen-Freeman & Long, 1991; Shavelson, Hubner & Stanton, 1976; Templin, 1995; Yule, Yanz & Tsuda, 1985), although some studies of self-efficacy mix it with these other self-phenomena (Huang & Chang, 1996; Mikulecky, Lloyd & Huang, 1996). Self-efficacy researchers specify five features that other self-phenomena researchers include only in part or not at all: (1) judgment of capabilities; (2) multiple dimensions; (3) contexts; (4) mastery-criterion; and (5)measurements taken before participants perform the task (Zimmerman, 1995). These are introduced below.

First, although self-efficacy is used as a judgment of capabilities (how well people believe they can do something), measures of other self-phenomena are often used as judgments of personal qualities (how well people feel about themselves). Second, self-efficacy researchers include multiple dimensions of research participants. Learners may believe they can introduce themselves orally, but they may not believe they can *write* a 50-word self-introduction. Other self-phenomena researchers do not always include multiple dimensions.

Third, self-efficacy researchers examine judgments of capabilities in various contexts. For example, learners may think they can introduce themselves in the context of a classroom of non-native English-speaking students, but they may think they cannot introduce themselves in a classroom of native English-speaking students. Although the task is the same, the context is different. Other self-phenomena researchers do not depend on context.

Fourth, while self-efficacy is based on mastery criteria, other self-phenomena are usually based on normative criteria. Self-efficacy researchers specify how well learners believe they can accomplish tasks. Other self-phenomena researchers usually compare what learners feel about themselves in comparison with what other learners feel about themselves—a method that includes no direct measurement of what learners think they can actually do.

Finally, self-efficacy researchers need to measure self-efficacy before learners actually perform their tasks. Other self-phenomena researchers measure the self-phenomenon before the task, after the task, or without performance of the task at all. If researchers measure their self-phenomena after the task, or do not require participants to perform the task at all, they can predict nothing.

Self-Efficacy Areas

Other self-phenomena researchers have also been largely unsuccessful in predicting human behavior, whereas self-efficacy researchers have been widely successful. Researchers have successfully studied self-efficacy in a variety of areas that include, but are not limited to, academic achievement (Lee & Bobko, 1994; Lent, Brown & Larkin, 1984; Wood & Locke, 1987; Zimmerman, 1995), career choice and development (Hackett, 1995; Matsui, Ikeda & Ohnishi, 1989; Matsui & Tsukamoto, 1991), and health (Schwarzer & Fuchs, 1995).

Psychology and management researchers have repeatedly predicted that students with high self-efficacy attain higher grade point averages than students with low self-efficacy. Similarly, as students finish school, those with high self-efficacy in career pursuits and personal health experience more success in their career pursuits and health than those with low self-efficacy.

Predicting L2 Learner Grades

In studies attempting to predict L2 learners' grades in ESL settings, applied linguists recommend exploring factors such as motivation, personality, attitudes, previous knowledge, and previous academic performance to predict academic achievement (Graham, 1987; Light, Xu & Mossop, 1987; Patkowski, 1991). Even though psychology and management researchers have predicted academic success from self-efficacy measurements, applied linguists have not explored self-efficacy measurements as a way to predict academic achievement in language classes.

Statement of Purpose

The purpose of this exploratory research is to see if high self-efficacy students will achieve significantly higher grades than low self-efficacy students in an L2 learning class.

Method

Participants

The 74 participants in this study were tenth grade Japanese nationals in an urban high school ranked eighth out of nine high schools in its area in Kanagawa Prefecture. Students were enrolled in English I, which focuses predominantly on grammar-translation with some oral/aural instruction. There were 35 females and 39 males, ranging in age from 15-17. Students were in two intact classes instructed by the same teacher. All students participated by filling out a research questionnaire (see Appendix) after they had taken their first semester midterm exam, but before they received the results of the exam. This was done so participants would have feedback about the course, but would not base their responses only on grades (Wood & Locke, 1987). No language proficiency scores were available for these students.

Instrument

Considering the low level of the participants' high school and teachers' observations that previous students had poor English skills, the self-efficacy instrument was created in Japanese so students could fully understand the questionnaire. Japanese native speakers (fluent in English) and a non-native Japanese speaker (native English speaker) created the questionnaire in Japanese then translated it into English for non-Japanese readers (see Appendix). Contact the author for the Japanese original.

The self-efficacy measurement was adapted from Locke and Latham's (1990, p. 348) instrument, a composite of self-efficacy magnitude and strength. Magnitude has been used to measure the differing levels that subjects believe they can perform in a given domain. In the domain of academic achievement in an L2 class, this study asks students whether or not they believe they can achieve the following grades in their English class: F-, F, D-, D, C-, C, B-, B, A-, A. It may seem that measuring ten levels of academic achievement (F- to A) is overkill. However, measuring one level (whether or not students believe they can achieve As) gives no information about the differences between students who only believe they can achieve other levels (Bs, Cs, etc.). The self-efficacy magnitude (see Appendix) shown in the left column, was obtained by asking students to answer yes or no if they could attain specific grades (F- to A). All data were entered into a *ClarisWorks 4.0* (ClarisWorks Corp., 1994) spreadsheet and analyzed using *Statview 4.5* (Abacus Concepts, 1995). The magnitude was then calculated by adding the total number of *yes* answers divided by the total number of items (10). Self-efficacy magnitude is the second most common self-efficacy measure in psychology and management research (Lee & Bobko, 1994). The most popular self-efficacy measure is self-efficacy *strength* (Bandura & Wood, 1989; Lee & Bobko, 1994; Matsui & Tsukamoto, 1991). People do not only differ in the levels of their efficacy beliefs (magnitude), but also differ in the strength of their efficacy beliefs:

> Weak efficacy beliefs are easily negated by disconfirming experiences, whereas people who have a tenacious belief in their capabilities will persevere in their efforts despite innumerable difficulties and obstacles. They are not easily overwhelmed by adversity (Bandura, 1997, p. 43).

The questionnaire in the Appendix shows strength in the right column: Students rated their degree of confidence (0-100%) in attaining each grade level (F- to A). Strength was then calculated by adding the scores and dividing them by the total number of items (10).

Rather than using magnitude and strength scores independent of each other, Lee & Bobko (1994) recommend combining magnitude and strength scores for stronger predictive validity. The *composite* is calculated by adding the raw self-efficacy strength for grade levels that students answered *yes* to. Self-efficacy strength for grades answered *no* to are excluded. Fewer researchers (Gist, Schwoerer & Rosen, 1989; McAuley, Wraith & Duncan, 1991) use the composite self-efficacy instrument.

Table 1 shows the results of one student's questionnaire. This student wrote that, yes (magnitude), she thought she could score an F- in the English class for a final grade. This student was 100% confident (strength) about this. This student thought she could not score an F in the class. The student's confidence in scoring an F was 50%. The student thought she could not score anything higher and had no confidence in attaining any higher grade. The researcher divided the number of *yes* scores (1) by the number of levels (10) for the student's magnitude score (.10). Then the researcher added all of the strength scores (.15 + .00 + .00, etc.) and divided by 10 for the student's strength score (.15). Finally, the researcher added all of the strength scores for *yes* answers (1.00 for F-). All strength scores for *no* answers (.50 for F, etc.) were excluded. This student's scores are the lowest scores in Table 2 for magnitude, strength, and composite. Although not observable from the data presented here, this student's final English grade was F (F=2).

Table 1: One Student's Magnitude, Strength, & Composite Scores

Grade (Yes/No)	Magnitude	Strength (.0-1.00 Confidence)	Composite (Strength of Yes)
F-	Yes	1.00	1.00
F	No	.50	.00
D-	No	.00	.00
D	No	.00	.00
C-	No	.00	.00
C	No	.00	.00
B-	No	.00	.00
B	No	.00	.00
A-	No	.00	.00
A	No	.00	.00
Scores	.10 (average)	.15 (average)	1.00 (sum)

Grades were determined by the teacher of the two classes by averaging grades for three semesters. These included grades for exams, assignments (in and out of class), and attendance and were represented on a scale of 1-10, the lowest score being 1 (F-) and the highest score being 10 (A).

Reliability of the Instrument

The reliability of the self-efficacy scores and grades were calculated using Cronbach's alpha and are reported in Table 2 below. The two subsections, magnitude and strength, and the composite of the questionnaire are .91, .98, and .96, respectively. The reliability of grades could not be determined because the necessary data were not available to the researcher.

During class the teacher passed out the questionnaire and gave students 10-15 minutes to fill it out. She suggested the students would probably answer *yes* with 100% confidence for the first question, since it is impossible to score lower than an F-. She did not recommend answers for any of the other questions.

After the students finished the questionnaires, the teacher collected them and sealed them in an envelope that she handed to the researcher after class. The teacher never saw the results of the questionnaires. At the end of the school year, the teacher gave her students' grades to the researcher.

Table 2: Descriptive Statistics for Self-Efficacy Scores and Grades

	Subtests			
Statistics	Magnitude	Strength	Composite	Grades
N	74.00	74.00	74.00	74.00
k	10.00	10.00	10.00	3.00
M	.53	.50	4.48	6.47
Mode	.50	.66	5.00	6.00
Median	.50	.49	4.30	6.00
Midpoint	.55	.55	5.30	5.50
Low-High	.10-1.0	.5-.96	1.0-9.6	1.0-10
Range	1.90	1.81	9.60	10.00
SD	.17	.16	1.60	1.98
Chronbach's Alpha	.91	.98	.96	*

*unavailable

Statistical Analysis

To analyze the data, descriptive statistics were calculated for the self-efficacy scores and grades (Table 2). The self-efficacy scores and grades have similar means, modes, medians, and midpoints. Differences were measured by a paired t-test, with an alpha level of .05.

Table 3: Low and High Self-Efficacy Students' Grades

	Groups	
Statistics	Low	High
N	37.00	37.00
k	3.00	3.00
M	5.89	7.05
Mode	6.00	7.00
Median	6.00	7.00
Midpoint	5.50	6.50
Low-High	1 - 10	3 - 10
Range	10.00	8.00
SD	1.89	1.92
SD squared	3.59	3.71

Table 4: Results of T-test Comparing Grades of Low & High Self-Efficacy Students

Groups	Mean Difference	df	t
Low, High	-1.16	36	-2.85*

*p < .05

Results

In order to compare the grades of low self-efficacy students with the grades of high self-efficacy students, the independent variable of this study was defined as the student's grade and the total number of participants, 74, was divided into halves. Those students who scored in the lower half on the self-efficacy composite were designated as the low self-efficacy group and students scoring in the upper half were designated as the high self-efficacy group. The descriptive statistics are given in Table 3.

Since both the low and high self-efficacy groups meet the assumptions of grouping, continuous data, normal distributions, and equal variance for a t-test, a one-tailed t-test was selected to compare group means (see Table 4).

As shown, the difference between the grades of low self-efficacy and high self-efficacy students was significant at $p < .05$.

Discussion

This pilot study suggests that high self-efficacy students achieve significantly higher grades than low self-efficacy students in an L2 classroom. From the beginning of the school year, low self-efficacy learners believe they cannot succeed academically and thus remain cut off from higher achievement throughout the year. This result is in agreement with self-efficacy research in psychology and management that shows low self-efficacy learners decrease attention, effort, persistence, and strategies for achieving, and they avoid challenging goals. While this researcher has observed that some students only exhibit low self-efficacy in language learning classes (e.g., they exhibit high self-efficacy in math, extracurricular activities, etc.), other students exhibit low appraisals of their capabilities across many of their school activities—a sign that these students may be in particular need of help.

Someone might argue that self-efficacy is just sound self-knowledge—people already know what they can and cannot do. But people do not always know what they can and cannot do (for more on the discordance between efficacy judgment and action, see Bandura, 1997, pp. 61-78). In dangerous situations where mistakes can be fatal, people kill themselves by overestimating their capabilities. However, in less dangerous situations, underestimating one's capabilities can lead to regret; "Educational opportunities forsaken, valued careers not pursued, interpersonal relationships not cultivated, risks not taken, and failures to exercise a stronger hand in shaping one's life course" (Bandura, 1997, p. 71).

Bandura (1995) cites research that shows four ways people can raise their self-efficacy. The first way is through *enactive mastery experience*. Learners need opportunities to experience success in L2 learning classrooms. Also, instead of measuring students' mastery using norm-referenced tests (NRTs) that only allow about 2% of the students to receive As, teachers should use criterion-referenced tests (CRTs) in their classrooms. Criterion-referenced tests allow 100% of the students to receive As and measure mastery of the coursework (Brown, 1996).

Second, learners can increase their self-efficacy through *vicarious experience*. When learners see their peers—whom they judge to be of similar L2 proficiency—fail, learners expect to fail. In contrast, learners who see their equals succeed believe they can succeed, too. Also, when Japanese teachers of English speak English, students believe that they can speak English, too.

Verbal persuasion is a third way learners can increase their self-efficacy. People can be persuaded verbally that they can succeed. Bandura (1995) explains,

> Successful efficacy builders do more than convey positive appraisals. In addition to raising people's beliefs in their capabilities, they structure situations for them in ways that bring success and avoid placing people in situations prematurely where they are likely to fail often. They encourage individuals to measure their success in terms of self-improvement rather than by triumphs over others. (p. 4)

Depending on what messages teachers send to their students, teachers can influence whether students have high or low self-efficacy.

Fourth, *physiological and affective states* affect learners' beliefs in their capabilities. Learners need to understand how to interpret feelings of arousal as positive, and learners need to be healthy. For example, before speaking in an L2, if students interpret their increased heartbeats, faster breathing, and higher perspiration as debilitating, they will lower their self-efficacy. Students with a positive interpretation will use the arousal to energize their performance. In addition, students need to get proper amounts of rest, eat a balanced diet, exercise regularly, etc. (For creating a self-efficacy syllabus in an EFL classroom, see Templin, in press.)

Although this study indicates that learners with high self-efficacy perform higher academically, it does not necessarily show that learners will successfully acquire the L2 studied. One difficulty with measuring L2 acquisition in Japanese academic institutions is that reliable and valid L2 proficiency measurements are rare. This researcher has advised and participated in language testing at the high school and university level, including administration of the Ministry of Education-endorsed *eiken* (tests produced by STEP, the Society for Testing English Proficiency). Reliable and valid testing is the exception rather than the norm (see articles in Brown & Yamashita, 1995), yet such measurements are needed so researchers can find out how much of the L2 learners actually acquire.

Also, using a composite of self-efficacy magnitude and strength scores is cumbersome to calculate. In this study, calculating strength alone seemed just as satisfactory as calculating a composite measure. Bandura (1997), says that calculating strength alone "provides essentially the same information and is easier and more convenient to calculate" (p. 44).

In future studies of academic achievement in L2 classrooms, it is suggested that researchers investigate self-efficacy instruments that measure the other dimensions of academic achievement such as concentration, memorization, and note-taking (Lee & Bobko, 1994; Wood & Locke, 1987).

Acknowledgments

The author thanks J.D. Brown, Dale T. Griffee, Nicholas O. Jungheim, Cynthia Lee, and Tamao Matsui for their comments regarding this manuscript. Correspondence should be addressed to Stephen A. Templin, Meio University, International Cultural Studies Department, 1220-1 Biimata, Nago-shi, Okinawa, Japan 905-0005. E-mail: steve@ics.meio-u.ac.jp. Work fax: 0980-52-4640.

Stephen A. Templin is the author of *Communicative Tool Box for Japanese Students* (Seido Language Institute). His articles have appeared in *JALT Journal*, *The Language Teacher*, *TESL Reporter*, and other publications.

References

Abacus Concepts. (1995). *StatView 4.5* (computer software). Berkeley, CA: Abacus Concepts.

Bandura, A. (1986). *Social foundations of thought and action* (pp. 390-453). Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-Efficacy in changing societies* (pp. 1-45).

Cambridge: Cambridge University Press.

Bandura, A. (1997). *Self-Efficacy: The exercise of control.* New York: W.H. Freeman.

Bandura, A. & Wood, R. (1989). Effects of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology,* 56, 805-814.

Brown, J.D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J.D. & Yamashita, S. (1995). *Language testing in Japan.* Tokyo: The Japan Association for Language Teaching.

ClarisWorks Corp. (1994). *ClarisWorks 4.0* [computer software]. Santa Clara, CA: Claris Corporation.

Earley, C.P. & Lituchi, T.R. (1991). Delineating goal and efficacy effects: A test of three models. *Journal of Applied Psychology,* 76, 81-98.

Ellis, R. (1990). *The study of second language acquisition,* p. 518. Oxford: Oxford University Press.

Gist, M.E. & Mitchell, T.R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review,* 17, 183-211.

Gist, M.E., Schwoerer, C. & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology,* 74, 884-891.

Graham, J.G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly,* 21, 505-521.

Griffee, D.T. (1997a). Using goals and feedback to improve student performance on vocabulary homework. *The Language Teacher,* 21(7), 19-25.

Griffee, D.T. (1997b). Validating a questionnaire on confidence in speaking English as a foreign language. *JALT Journal,* 19, 177-197.

Griffee, D.T. & Templin, S.A. (1998). Goal-setting affects task performance. In B. Visgatis (Ed.), *On JALT '97: Trends & Transitions* (pp. 21-26). Tokyo: Japan Association for Language Teaching.

Hackett, G. (1995). Self-efficacy in career choice and development. In A. Bandura (Ed.), *Self-Efficacy in changing societies* (pp. 232-258). Cambridge: Cambridge University Press.

Heyde, A. (1979). *The relationship between self-esteem and the oral production of a second language.* Unpublished Ph.D. Dissertation, University of Michigan, Ann Arbor.

Huang, S.C. & Chang, S.F. (1996). *Self-efficacy of English as a Second Language learner: An example of four learners* (Report No. FL 023 879). Washington, DC: Clearinghouse on Languages and Linguistics. (ERIC Document Reproduction Service No. ED 396 536).

Larsen-Freeman, D. & Long, M.H. (1991). *An introduction to second language acquisition research,* p. 184. London: Longman.

Lee, C. & Bobko, P. (1994). Self-efficacy beliefs: Comparisons of five measures. *Journal of Applied Psychology,* 79, 364-369.

Lent, R., W. Brown, S.D. & Larkin, K.C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psy-*

chology, 31, 356-362.

Light, R.L., Xu, M. & Mossop, J. (1987). English proficiency and academic performance of international students. *TESOL Quarterly,* 21, 251-261.

Locke, E.A. & Latham, G.P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Matsui, T., Ikeda, H. & Ohnishi, R. (1989). Relations of sex-typed socializations to career self-efficacy expectations of college students. *Journal of Vocational Behavior,* 35, 1-16.

Matsui, T. & Tsukamoto, S.I. (1991). Relation between career self-efficacy measures based on occupational titles and Holland codes and model environments: A methodological contribution. *Journal of Vocational Behavior,* 38, 78-91.

McAuley, E., Wraith, S. & Duncan, T.E. (1991). Self-Efficacy, Perceptions of Success, and Intrinsic Motivation for Exercise. *Journal of Applied Social Psychology,* 21 (2), 139-155.

Mikulecky, L., Lloyd, P. & Huang, S.C. (1996). Adult and ESL literacy learning self-efficacy questionnaire (Report No. FL 023 879). Washington, DC: Clearinghouse on Languages and Linguistics. (ERIC Document Reproduction Service No. ED 396 536).

Patkowski, M.S. (1991). Basic skills tests and academic success of ESL college students. *TESOL Quarterly,* 25, 735-738.

Schwarzer, R. & Fuchs, R. (1995). Changing risk behaviors and adopting health behaviors: The role of self-efficacy beliefs. In A. Bandura (Ed.), *Self-Efficacy in changing societies* (pp. 259-288). Cambridge: Cambridge University Press.

Shavelson, R., Hubner, J. & Stanton, G. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research,* 46, 407-441.

Templin, S.A. (1995). Goal-setting to raise speaking self-confidence. *JALT Journal,* 17, 269-273.

Templin, S.A. (in press). Self-efficacy syllabus. *The Language Teacher,* 22 (4).

Wood, R.E. & Locke, E.A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement,* 47, 1013-1024.

Yule, G., Yanz, J.L. & Tsuda, A. (1985). Investigating aspects of the language learner's confidence: An application of the theory of signal detection. *Language Learning,* 35, 473-488.

Zimmerman, B.J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-Efficacy in changing societies* (pp. 202-231). Cambridge: Cambridge University Press.

(Received February 6, 1998; revised October 15, 1998)

Appendix: Self-Efficacy Questionnaire (English Version)

___Year ___ Class ____ID Male_ Female_ Name______________________

(Your teacher will not look at this, and your answers will not affect your grades.)

In this class (for your final grade),

Do you think you can score an F-?
Yes__ No__
Do you think you can score an F?
Yes__ No__
Do you think you can score a D-?
Yes__ No__
Do you think you can score a D?
Yes__ No__
Do you think you can score a C-?
Yes__ No__

Do you think you can score a C?
Yes__ No__
Do you think you can score a B-?
Yes__ No__
Do you think you can score a B?
Yes__ No__
Do you think you can score an A-?
Yes__ No__
Do you think you can score an A?
Yes__ No__

How much confidence do you have that—

You can score an F-?
(0% - 100%)_______
You can score an F?
(0% - 100%) _______
You can score a D-?
(0% - 100%) _______
You can score a D?
(0% - 100%) _______
You can score a C-?
(0% - 100%) _______

You can score a C?
(0% - 100%) _______
You can score a B-?
(0% - 100%) _______
You can score a B?
(0% - 100%) _______
You can score an A-?
(0% - 100%) _______
You can score an A?
(0% - 100%) _______

Note: The original Japanese questionnaire can be obtained by contacting the author.

A Myth of Influence: Japanese University Entrance Exams and Their Effect on Junior and Senior High School Reading Pedagogy

Bern Mulvey
Fukui University

In discussions regarding the negative aspects of exam "washback effect," one example that is invariably mentioned is the exam-pedagogy relationship ostensibly to be found in Japan. Indeed, it is the supposedly powerful influence of the various university exams on junior and senior high school classroom pedagogy and textbook content in Japan that allegedly both perpetuates inadequate teaching methodologies and frustrates all attempts at reform. This paper examines the large body of research that calls into question this traditional conception of a causal relationship between the entrance exams and junior and senior high school foreign language reading pedagogy and textbook content, and hypothesizes as to the possible non-exam-related motivations for the continued use in Japan of seemingly ineffective foreign language reading pedagogy.

本研究では、試験が教授法に及ぼす「望ましくない波及効果」について論じる。日本の大学入学試験が、中学校・高等学校における英語教育の発展にとって障害となっていることが、多くの研究者によって指摘されている。本研究では、中学校・高等学校での英語読解指導方法や検定済教科書を調査し、指摘されているような大学入学試験の波及効果について、その実態の分析を行った。この分析結果をもとに、教授法や教科書のあり方について提案を行う。

This paper asserts a position that many at first glance will consider untenable—that the influence of the various university exams (i.e., both the national entrance exam and the various independently generated and separately administered individual college or faculty exams) on junior and senior high school foreign language pedagogy in Japan has been exaggerated. Furthermore, this paper makes another equally controversial claim—that the content of these exams can neither explain nor justify the extreme inadequacy of the methodology currently used to teach English reading skills in the overwhelming majority of Japan's junior and senior high schools.

JALT Journal, Vol. 21, No. 1, May, 1999

The received arguments in place against these positions are formidable. Almost all the studies referred to in this paper agree that there are serious problems with English education in Japan; however, the literature to date *never* fails to identify the ostensibly powerful, and allegedly damaging, influence of the entrance exams as a primary cause of these problems. Indeed, advocates of reform (see Brown, 1993; Brown, 1995; Brown & Yamashita, 1995a & b; Ishizuka, 1997; Rohlen, 1983; Shimaoka & Yashiro, 1990; Sturman, 1989; Vanderford, 1997) focus almost exclusively on the supposedly inhibitive effect of these exams in their current form on attempts to improve junior and senior high school teaching methodology and textbook content. Other observers (such as Cutts, 1997; Frost, 1991; and Tsukada, 1991) note in detail the "big business" aspects of the service industry (the so-called "*juku-yobiko*" system) that has grown up around preparing students for these exams, and they discuss at length the implications of the powerful influence that the existence of this industry suggests. Finally, critics such as Hards (1998) and McNabb (1996) take an even more extreme position, holding that the exams are solely responsible for a host of assorted educational problems, and arguing further that they must be done away with entirely.

A key term that many of these writers use in making these observations is "washback effect," in this case used to refer to the supposed cause-and-effect nature of entrance examinations' influence on junior and senior high school teaching methodology. The content of these exams, we are told, dictates to a great extent how and what students will be taught up until they graduate from high school. As Brown says in an interview published in *The Language Teacher* (Leonard, 1998),

> It definitely goes on. Basically, teachers teach to prepare for particular tests. The same is true for the yobiko and juku [cram schools]. In fact, these schools gain customers by having a proven track record with certain exams. There is a really high anxiety level involved with these exams—studying for them and getting ready for them (p. 26).

Many writers agree with this position. Sturman (1989), for instance, writes, "the final aims of schools is to prepare students for entrance examinations" (p. 76). Tsukada (1991), among others, delineates at length the ways in which this influence has "undesirable effects on curriculum, on foreign language instruction, on family life, and on children's emotional, physical, and intellectual development" (p. 178) (see also, Frost, 1991, for similar commentary).

Furthermore, both this influence and the so-called "language testing hysteria" (Brown, 1993, 1995) that it engenders are used to support a further assertion, that merely by instituting changes to (or even eliminating) the exams, one will achieve beneficial changes in the educational system as a whole. Indeed, it is their belief in the strength of this cause-and-effect relationship between exam contents and classroom pedagogy in Japan that enables Vanderford (1997) to assert confidently that if the entrance exams but contained, "a reliable and valid test of oral English, I believe teachers and students [would] follow suit by teaching and studying English in a more communicative way" (p. 23), or allows Brown (1995) to state,

> Teachers should also recognize the relationship between the item types used on university entrance examinations and the pedagogical choices that they make in their classrooms. In 1993 and 1994, the private universities predominately used discrete-point receptive items. This means that in effect they were endorsing a discrete-point receptive view of language teaching (p. 97).

and later,

> Japanese universities should begin to change their examinations in similar ways so that their washback effect can become a positive and progressive force for change in language teaching in Japan (p. 98).

Again this implies that the contents of these exams are somehow responsible for the pedagogical practices and textbook content in use at the junior and senior high school level throughout Japan.

Impetus for Writing

The impetus for writing this paper arose out of the author's first-hand experience with the entrance exam process here in Japan, including three years as a member of the committee for making and grading the English entrance exams (*Eigoka Nyuugaku Shiken I-Inkai*), the committee for deciding the form and content of all entrance exams at the university (*Nyuugakusha Sembatsu Houhou I-Inkai*), and the committee for making the final decisions as to who is to be accepted into the university (*Nyuugaku Shiken I-Inkai*). During this period, the author noted that over 50% of the would-be English and/or Education students did poorly on the English portion of the entrance exam (in this case, "poorly" refers to those scoring less than 60% correct on the test). However, only 20% of the students applying for entry into either of these programs were turned away. This meant that about 30% of the incoming Education and English majors were accepted into the freshman class despite doing poorly on these exams.

Furthermore, although students generally answered grammar questions correctly, questions focusing on listening and reading comprehension skills were either answered incorrectly or were skipped entirely. Certainly, considering the nature and pervasiveness of the stereotype that "Japanese know grammar, reading and writing but can't speak" (see Hards, 1998, and Shimaoka & Yashiro, 1990, for instance), one is not surprised to learn that Japanese students did poorly in listening. However, their not being able to understand reading passages with an average Gunning's Fog Index rating of 11.600[1] after 6 years of English education was another matter. Where were the fruits of the intensive (an average of 3 hours a week in junior high and 6 hours a week in senior high, not including time spent at *juku-yobiko*) reading and grammar-centered "test preparation" that these students supposedly had undergone?

In order to answer the above question, this author examined 51 studies containing analysis of the methods used and the skills taught in English reading classes at the junior and senior high school level. Since many of these studies are written in Japanese, this report will mark the first time that much of this research is made available to non-Japanese readers. The results of these studies were then compared to the reading skills areas evaluated by the various university entrance exams. The results were indeed surprising. There seemed to be little direct evidence of a causal relationship between entrance exam content and either textbook contents or junior and senior high school English reading pedagogy, at least with regards to the teaching of reading skills. This is in direct contradiction to the monolithic block of critical commentary cited above.

This paper presents the results of these studies and analyzes the areas of weakness in Japanese readers of English that these studies have pointed out, and the possible reasons for these weaknesses. Finally, it hypothesizes as to the possible motivations for the continued use in Japan of reading methodology that does not assist, and may in fact impede, the acquisition of English reading skills.

Review of Research

Far from the test "cart" pulling the educational "horse," the contents of the various Japanese university entrance exams seem to have had negligible effect on reading textbook content, reading pedagogy, and/or improving overall student capabilities. Reading skills sections of university entrance exams have been analyzed by Brown (1995), Law (1994), Kimura & Visgatis (1996), and Pai (1996), among others, with the following conclusions:

1) The reading passages used therein are almost without exception adult level, well-written, grammatically and stylistically correct (see Brown,

1995, pp. 96-97; Law, 1994, p. 96; Kimura & Visgatis, 1996, pp. 86-92; Pai, 1996, p. 153)

2) Contextualized, task-based questions (i.e., not just translation or narrow "discrete-item" questions) make up a large portion of these exams, requiring examinees to have the ability to summarize and/or explain difficult areas in the reading passages (see Brown, 1995, pp. 94-95; Law, 1994, p. 96; Kimura & Visgatis, 1996, pp. 86-92; Pai, 1996, p. 153).

In other words, in order to be prepared for these exams, university-bound high school students would need both to have learned "to read relatively difficult university level passages with good comprehension" (Brown, 1995, p. 96), and to have developed the "rapid structural and lexical recognition skills" (Law, 1994, p. 98) necessary to answer the "integrative" (i.e., reading comprehension) questions that come with such passages (see also Kimura & Visgatis, 1996, pp. 86-92; Pai, 1996, p. 153).

Certainly, mastering the above skills would not be an easy proposition even if the six years and almost one thousand hours of language instruction that college-bound Japanese students typically receive was really the reading- and grammar-centered test preparation that it is held to be. However, analyses of teaching materials and observational studies of classroom methodology conducted by Gorsuch (1998); Hino, (1988); Jannuzi, (1994); Kimura & Visgatis, (1996); Kitao & Kitao (1989, 1995); Kitao, Kitao, Nozawa & Yamamoto (1985); Kitao and Yoshida, (1985); Law, (1994); Mulvey, (1998); Nishijima, (1995); Pai, (1996); Saeki, (1992); Takefuta, (1982); Tanaka, (1985); H. Yoshida, (1985); S. Yoshida, (1985); and Yoshida & Kitao, (1986), among others, raise serious questions about the nature and content of the supposed "test preparation" that Japanese students are being made to undergo.

First, there appears to be little correlation between the reading materials used at the junior and senior high school level and the contents of the various university entrance exams. Kimura & Visgatis (1996), for instance, conducted both Flesch-Kincaid and Gunning-Fog grade level analyses of the contents of several textbooks and entrance examinations, finding the reading difficulty of the entrance exam materials to be:

> three or more grade levels above the materials they have been exposed to. . . . This is even more striking after considering that students using textbooks are free to read the passages at home, consult reference works (i.e. dictionaries), and are not subject to the rigorous time constraints found under examination conditions (p. 90).

Pai (1996) comes to similar conclusions, noting that many junior and senior high school textbook reading passages are "full of grammar, spelling, syntactical and stylistic mistakes," and commenting that, outside of those attending college-prep classes at elite high schools (which also use old entrance exams), most Japanese students will receive "no exposure to adult level, well-written, and error-free reading passages before sitting for an university entrance exam" (p. 153; see also Law 1994). Furthermore, Kimura & Visgatis (1996) also assert the following,

> [I]t might be assumed that students are faced with progressively more difficult reading materials as they proceed through the high school curriculum, thus being amply prepared for the difficult reading passages found on entrance examinations. Unfortunately, this is not borne out by the textbook materials. Examination of the difficulty patterns of textbook reading passages shows that the highest average Flesch-Kincaid reading level does not appear in the last third of any of the textbooks, and only two of the textbooks have the most difficult Gunning-Fog result in the final third. If the chapters of the books are used sequentially, students will not be facing the most difficult passages at the end of their high school tenure (p. 90).

The citations above raise two important considerations. If the purpose of secondary-level education in Japan is to prepare students for the university entrance examinations, one would expect textbook content to reflect what is actually on these exams. Furthermore, one would expect textbooks to be designed with progressively increasing difficulty levels in order to slowly acclimate students to the skill-levels needed to succeed on these exams. However, the textbooks are not designed this way, and especially considering the three grade-level difference between textbook and test contents, one is forced to at least question the nature of the "test" preparation that is going on in these classrooms. In other words, where is the exam "washback effect" in an educational system where the contents of the textbooks bear so little relevance to the tests themselves?

Moreover, while effective classroom methodology could go a long way toward making up for any deficiencies in textbook content, there is much evidence to suggest that the methodology being used in Japan's junior and senior high schools is not effective. As noted above, the reading passages on entrance exams are generally native-speaker level in complexity, with the relevant questions that the students must answer most often integrative/comprehension in nature, i.e., ones that demand advanced structural and lexical recognition skills. Regarding the teaching of such skills to ESL/EFL students, while the issues involved remain somewhat controversial (see Gu, 1996, pp. 11-12), a majority of researchers, including Carrell (1987), Carrell & Eisterhold (1983), Grabe

(1991), Rumelhart (1977, 1980), and Sanford & Garrod (1981), have long argued that "both top-down and bottom-up strategies operating interactively" are necessary for students to be successful (Carrell, 1987, p. 24). Hence, an effective methodology, especially one with the averred goal of preparing students to read and respond to the native speaker-level passages used on entrance exams, would seemingly be one that attempted to provide students with both bottom-up and top-down strategies. These include strategies for analyzing the words and sentences in the text itself (such as guessing from context or skimming) and for making use of students' own experiences (i.e. their cultural and linguistic background knowledge) to illuminate those areas of meaning left indecipherable by bottom-up processing alone.

However, studies by Gorsuch (1998), Hino (1988); Jannuzi (1994); Kitao & Kitao (1995), Kitao et al. (1985), Kitao and Yoshida (1985), Law, (1994, 1995), Mulvey, (1998), Nishijima (1995), Takefuta (1982), Tanaka (1985), H. Yoshida (1985), S. Yoshida (1985), Yoshida & Kitao (1986), and Yukawa (1994), among others, suggest that the reading pedagogy employed in most Japanese schools is severely deficient in its presentation of both bottom-up and top-down approaches. While the methodology used in Japanese high school classrooms is certainly not identical in all cases, the above studies have identified the following elements as common to the methodology at most schools. First, despite research questioning its effectiveness (see Kitao et al., 1985; Kitao & Kitao, 1995; Kobayashi, 1975; Tanaka, 1985), teacher led and dominated line-by-line translation remains the preferred teaching methodology most students will encounter in the 6 years leading up to their entrance into college (Hino, 1988; Jannuzi, 1994; Kitao et al., 1985; Mulvey, 1998; Robb & Susser, 1989). Second, content-based questions, such as the kind featured on most entrance exams, are rarely used as teaching tools in most junior and senior high school classes, and if they are used (such as at elite college-prep schools where old exams are used to supplement the textbooks), students are rarely given the opportunity to individually negotiate meanings in a particular passage. (Kitao, Kitao, Nozawa, & Yamamoto, 1985). Instead, teachers in many cases literally dictate the correct answers in Japanese to the students, whose role it is to take notes to be regurgitated verbatim on later tests (Gorsuch, 1998, pp. 22-23; Kitao & Kitao, 1995, pp. 147-167; Mulvey, 1998; Saeki, 1992, pp. 18-19). Indeed, in a written survey given in Japanese to incoming freshmen (312 students) at Fukui University over a period of 2 years, 68% said that they had spent less than 2 hours a month reading English passages (in class or out) in junior and senior high school, and a full 72% characterized what "reading" they had done as translation exercises (Mulvey, 1998). Furthermore, an amazing 92% reported having had neither an opportunity to discuss nor to analyze independently the thematic contents of the passages they did read, stating instead that they were merely dictated answers that they were then expected to memorize for later tests.

One result of the above-described methodology is that, outside of the grammar emphasis, standard reading and comprehension strategies are just not taught at most high schools: skimming and/or guessing from context strategies are neither encouraged nor explained (Kitao, 1979; Kitao, Yoshida & Yoshida, 1986; Kitao & Kitao, 1995, pp. 147-167; Tanaka, 1985); word relationships (such as between synonyms and/or antonyms) are not taught (Kitao, Broderick, Fujiwara, Kitao, & Sackett, 1985; Kitao, Yoshida & Yoshida, 1986), and a significant percentage of students never even learn to use a dictionary effectively by themselves (Kitao et al., 1985; Kitao, Yoshida & Yoshida, 1986); limited English reading practice in junior and senior high school leaves students with difficulties recognizing Roman script (Weaver, 1980) and English sentence word order (Kitao, 1979; Kitao, Yoshida & Yoshida, 1986); and finally, English vocabulary (Kitao & Kitao, 1995, pp. 147-167; Kitao et al., 1985) and reading speed (Yoshida, S., 1985; Yoshida & Kitao, 1986)—even after six years and almost 1,000 hours of study—remain completely inadequate to allow reading comprehension of anything approaching authentic English texts.

Top-down processing strategies such as scripts, schemes, and the use of students' background knowledge or experiences also are not addressed. For instance, students are not taught culturally specific, preferred organizational differences (Kitao & Kitao, 1989, 1995). These include differing methods of topical progression and/or rhetorical organization as described in work by Hinds 1983, 1990; Kobayashi, 1984; Mulvey, 1992; Ricento, 1987; and Yutani, 1977, knowledge of which might enable students to better anticipate the topical progression in a particular work. Moreover, most high school teachers are not even aware of the 30+ years of relevant research (Kawasaki, 1998). Strategies for relating pieces of information as a way of increasing reading retention capacity have not found their way into most high school curriculums (Takahashi & Takahashi, 1984). Due to the superficial content of most "comparative cultures" education in Japan, students often never receive the cultural background knowledge necessary to make key connections and recognize implied meanings (Kitao & Kitao, 1989, 1995). Finally, even in many *Japanese* literature classes, with their long tradition of non-text-centered and non-analytical pedagogy (Hatano, 1993; Inoue, 1993; Sakamoto, 1995, p. 261), students rarely practice the kind of "reading for comprehension" skills demanded on the English reading sections of the entrance exams, resulting in students who are unaccustomed to analyzing passages in this way in their own language being asked to do so (for the entrance exams) in another (Gorsuch, 1998, p. 23; Kitao & Kitao, 1989, 1995).

In other words, researchers have shown that few Japanese students receive adequate bottom-up preparation in reading. Furthermore, even those who do have been found to have extreme difficulties reading authentic texts, both because of their lack of exposure to such texts and because they have not been exposed to the top-down strategies necessary to fully appreciate them. And again, as the ability both to understand and to respond to authentic English texts is one of the ostensible goals of the six years of preparation that Japanese students receive before sitting for the exams, the deficiencies in both top-down or bottom-up preparation that have been delineated throughout this paper must perforce call into question the nature of the relationship between exam content and the "test-centered reading preparation" that Japanese students are supposedly receiving. In other words, where in all the above-documented lack of reading preparation is there evidence of a causal relationship between test and pedagogy in Japan as described by Brown, (1993); Brown, (1995); Brown & Yamashita, (1995a & b); Ishizuka, (1997); McNabb, (1996); Rohlen, (1983); Shimaoka & Yashiro, (1990); and Vanderford, (1997)? Given that it generally produces—and indeed seems almost designed to produce—students with limited context-recognition skills, poor vocabularies, inadequate rhetorical/ schematic preparation, and deficient cultural background knowledge, i.e., just the areas that a truly "test-centered reading curriculum" would seemingly emphasize, it seems safe to say that both the nature and the extent of the exam's "washback effect" on the educational system in Japan have been exaggerated. At the very least, the above discussion suggests that the relationship between test content and the perpetuation of current pedagogical practices is actually extremely complex and may involve a variety of contributing factors.

While they are careful to place the majority of the blame on exam influence, other researchers have recently begun to search for additional, possibly contributing, factors. For instance, Gorsuch (1998), Hino (1988), Jannuzi (1994), Kitao & Kitao (1995), Kitao et al. (1985), Law (1994, 1995), and Yukawa (1994) suggest that teaching grammar in English reading classes, including the intricacies of *Japanese* grammar, are important classroom goals. Jannuzi (1994), for example, relates this about the large number of reading-centered classes he either observed or participated in during the four years he spent teaching in Japanese high schools:

> [T]ranslation was almost always from English into Japanese. If students did undertake translation, it was limited to the translation of sentences disconnected from longer discourse in order to practice grammar points. Students did not translate authentic texts (p. 122).

Hino (1988), Law (1994, 1995), and Gorsuch (1998) report similar findings. Hino writes that the teacher's role in the classroom is to "provide a model translation, and to correct the student's translation" (p. 46), to which Law (1995) adds, "the focus of attention is only initially on the codes of the foreign language; most of the productive energy of the method is directed towards the recoded Japanese version" (p. 216). Gorsuch (1998), finally, writes that the classroom methodology she observed,

> appeared to the researcher more as lessons in Japanese than in English. On one hand, these sequences served to help teachers focus students' attention on grammatical differences between English and Japanese. On the other hand, the teachers focused on helping students to think about and create meaningful Japanese, rather than meaningful English (p. 20).

Even more interestingly, Gorsuch (1998) relates that both teachers she observed, when interviewed, admitted that helping students "learn Japanese" is an important part of what they are attempting to achieve through their English reading classes (p. 23), again supporting the conclusions of the other researchers. Indeed, if the above observations are accurate, it would seem that teaching proper Japanese grammar is an important supplementary goal in at least some English classrooms, providing one additional explanation for the oft-observed heavy reliance in this country on line-by-line translation into Japanese as a foreign language instructional tool.

Additional ulterior motives for the continued use of the present methodology have also been suggested. Hino (1988), for instance, asserts that this methodology builds mental discipline in the students. Law (1994) interprets its continued utilization as almost reflecting a xenophobic element in the Japanese national character, arguing that it is a symbol of a Japan's "refusal of direct engagement" with other languages and its unwillingness to deal with the "codes" of a foreign culture without "recoding" them into Japanese (p. 97). Gorsuch (1998) suggests that the need to maintain "control" in the classroom is a prominent motivational force, writing that this pedagogy "affords teachers powerful control over students' language learning activities," and noting, "students were required to translate at nearly every juncture, and their translations were checked, and controlled, by the teachers in and out of class" (p. 27).

Finally, there is one further possibility. Judging by this author's three years of experience as a Literature instructor at the only teacher training program in the prefecture, many would-be Japanese teachers of English appear to receive little exposure to or training in reading pedagogy outside of that described in the preceding sections above. In other words, could teacher ignorance of possible pedagogical alternatives be an additional contributing factor in the perpetuation of current methodological practices? After all, people have been criticizing English pedagogy in Japan for the same reasons for over 100 years (see Mantanle, 1996), from a time preceding the university entrance exams in their current manifestation.

Certainly, a much broader study would be necessary to establish any of these conclusions as definitive. However, it should be clear from the above hypotheses that other researchers are at least beginning to question the motives behind the pedagogical practices in use at Japanese schools. Indeed, given the apparent irrelevancy of current methodology in assisting students in passing at least the reading sections of the entrance exams, it seems possible to argue that there is at least the chance of strong motivational forces and situational requirements operating here outside of mere "test preparation," ones that have not been fully studied but which may be significant nonetheless.

Conclusions and Final Comments

In arguing that the washback effect of the university entrance exams on reading pedagogy has been exaggerated, this author wishes to make clear that he is neither overlooking nor discounting the integral and often negative impact of the exams on the Japanese economy, social and educational system, and family. That there is an "exam hysteria" (Brown 1993, 1995) is self-evident; that a lot of time and especially money is invested in this multi-billion dollar industry is undeniable (Frost, 1991); that the effect on Japanese family life and, in particular, the effect on high school students caught in "exam hell" can be and often is devastating is also unarguable (Tsukada, 1991).

Less apparent, however, is the connection between the reading pedagogy in practice at most junior and senior high schools in Japan and the entrance exams that have supposedly necessitated it. Native-speaker level reading passages and related comprehension and analytical questions are on the entrance exams: Where is the preparation for handling these types of passages and questions? Furthermore, entrance exam questions seem to be becoming progressively more analysis- and comprehension-centered (Brown & Yamashita, 1995a & b; Law, 1994, 1995). At the same time, however, the overall ability of Japanese students to handle such questions or to read authentic English passages seems to actually be *decreasing* (Ishizuka, 1997; Nishijima, 1995; Saeki, 1992, p. 28). Study after study discussed in this paper supports these latter findings. In addition, they point out the probable explanations for this phenomenon: poor bottom-up and top-down preparation, little to no exposure to extensive reading with authentic English texts, and a lack of opportunities to independently negotiate textual meanings or to attempt to master comprehension questions on their own. Where, then, is the "washback effect" on pedagogy that these exams are supposed to produce?

Is all this simply a problem of the entrance exams being too difficult, as suggested by some writers (see Brown 1993, 1995; Brown & Yamashita, 1995a & b, and Kimura & Vigatis, 1996)? This is a complex question. That the reading sections of many of these exams *are* too difficult for most Japanese students is obvious. Less obvious, however, is whether the skill levels demanded by the exams represent excessive or unreasonable expectations for students with six years and almost one thousand hours of intensive, supposedly reading and grammar-centered, academic preparation. In addition, what is "normal" for the rate of acquisition of L2 reading skills in a non-European EFL population is something which is not established, since little research has been done in this area. For example, studies conducted by Cummins (1981) and Ekstrand (1976, 1978) deal only with children in an ESL environment; Grinder, Otomo & Toyota (1962) looks at the acquisition of EFL *listening* skills in elementary school-age Japanese children; and Collier (1987) and Kuroiwa (1997), the two most relevant studies found and ones whose findings seem to support the argument that Japanese students should be much better prepared than they are, look only at the ESL acquisition rates of students in relation to their length of stay in the country where the L2 is spoken. Hence, even these latter studies are not really applicable to the EFL situation.

Does this lack of relevant research protect Japanese schools from the charge that they are not doing all they can to give students the reading skills necessary to succeed on the entrance exams? Hardly. As the research cited in this paper illustrates, current methods of teaching EFL reading in Japan are grossly inadequate and result in a large number of students who have difficulty understanding texts written in English. These findings of inadequacy are further supported by a comparison of average TOEFL scores between Japan and other Asian countries. Although such a comparison certainly cannot be taken as definitive in itself, the results in this case are suggestive. Despite the fact that Japan spends far more on foreign language education, despite the fact that Japanese students receive on average far more hours of English instruction per week, and despite the equivalent levels of difficulty in moving from the L1 to the L2, Korean, Taiwanese, Chinese, and Thai students all have significantly higher average TOEFL reading scores than their Japanese counterparts: 499 for the Japanese, compared with 519/520/556/520 respectively for the other groups (Ishizuka, 1997; *Keizai doyukai*, 1998, pp. 206-213; Saeki, 1992, p. 28). Moreover, the traditional rebuttal to such statistics—that only the elite students from the other countries listed take the TOEFL—does not hold up to close examination. Although more Japanese do take the exams, the percentage of the total Japanese population taking the exams is actually *lower* than that of Korea and Taiwan.[2] Hence, it could be just as easily argued that it is the Japanese educational elite that are taking and doing poorly on the exams in high numbers.

Furthermore, it should also be noted that the average TOEFL reading scores of Japanese students have continued to *decrease* steadily over the last 20 years, ironically, while speaking scores have gone up (see Ishizuka, 1997). This is a failure that is occurring despite the presence of adult native speaker-level reading passages on the college entrance exams, the increasing use on the exams of comprehension questions demanding advanced structural and lexical recognition skills, and the reading-centered teaching methodology that this usage ostensibly should have engendered. Again, where is the evidence in this gradual decline of reading skills of either an exam "washback" effect or six years of supposedly intensive "grammar- and reading-centered" test preparation?

Finally, this author noted earlier in this paper that, in his experience, would-be students regularly do poorly on the entrance exams and yet are still accepted into college. Is this experience an aberration? Several commentators (Leonard, 1998; Vanderford, 1997, p. 19) have noted the critical role of recommendations and/or athletic scholarships in the post-secondary school admissions of up to 30% of Japanese students. Furthermore, consider the following. In America, traditionally considered a country with lax admissions standards, 70% of students go on to enter post-secondary/tertiary schools (i.e., either two-year or four-year colleges). In Japan, a country long noted for the strictness of its admissions policies, an almost equal 69% go on to successfully enter post-secondary/tertiary schools *(Keizai doyukai*, 1998, p. 216). In other words, despite apparently low average skill levels when compared to the demands of the various exams, most Japanese students do manage to go on to post-secondary schools.

In short, the assumption of many of the writers referred to at the beginning of this paper, i.e., the importance of these entrance exams and their supposed "washback effect" on pedagogy in Japan, is actually a somewhat controversial premise worthy of a more open and critical debate. Indeed, as the overall pool of Japanese students attempting to get into post-secondary schools continues to decrease due to a declining birthrate and other demographic forces, it stands to reason that post-secondary programs will be forced to compete more energetically in order to maintain enrollment at levels sufficient to ensure their economic viability, including, perhaps, a continued relaxation of admission standards. With such motivational forces and situational requirements in mind, it seems clear that the importance of the entrance exams and the relevancy of the preparation that students are receiving for them will become an increasingly controversial issue in the foreseeable future. It is hoped that the research discussed in this paper will help further debate on this issue.

Bern Mulvey is a professor of American Literature at Fukui University.

Notes

1. This indicates a readability level approximately equivalent to the U.S. mid-third year level in high school. The author recognizes the limitations of such indexes as measuring devices of passage complexity. However, their use as a means of providing general indications of passage difficulty is long established (see Crystal, 1987; Richards, Platt & Weber, 1985).
2. Based on 1994 statistics [author's note]. See, for instance, Ishizuka (1997).

References

Brown, J.D. (1993). Language testing hysteria in Japan? *The Language Teacher*, 17 (12), 41-43.

Brown, J.D. (1995). English language entrance examinations in Japan: Problems and solutions. *JALT 95 Conference Proceedings*, 273-283.

Brown, J. & Yamashita, S. (1995a). English language tests at Japanese universities: What do we know about them? *JALT Journal*, 17 (1), 7-30.

Brown, J. & Yamashita, S. (1995b). English language entrance examinations at Japanese universities: 1993 and 1994. In Brown & Yamashita (Eds.), *Language Teaching in Japan* (pp. 86-100). Tokyo: JALT.

Carrell, P.L. (1987). A view of written text as communicative interaction: Implications for reading in a second language. In Joanne Devine, Patricia L. Carrell, and David E. Eskey (Eds.). *Research in reading English as a second language* (pp. 21-36). TESOL.

Carrell, P.L. & Eisterhold, J.C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17 (3), 553-573.

Collier, V.P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, 21 (4), 617-641.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge: UP.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 2, 132-149.

Cutts, R.L. (1997). *An empire of schools: Japan's universities and the molding of a national power elite*. London: M.E. Sharpe.

Ekstrand, L. (1976). Age and length of residence as variables related to the adjustment of migrant children, with special reference to second language learning. In G. Nickel (Ed.). *Proceedings of the Fourth International Congress of Applied Linguistics* (Vol. 3, pp. 179-197). Stuttgart: Hochschulverlag.

Ekstrand, L. (1978). English without a book revisited: The effect of age on second language acquisition in a formal setting. *Didakometry, No. 60*. Malmö, Sweden: School of Education, Department of Educational and Psychological Research.

Frost, P. (1991). Examination hell. In E.R. Beauchamp (Ed.), *Windows on Japanese education* (pp. 291-305). New York: Greenwood Press.

Gorsuch, G.J. (1998). Yakudoku EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20 (1), 6-32.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25 (3), 375-406.

Grinder, R., Otomo, A. & Toyota, W. (1962). Comparisons between second, third, and fourth grade children in the audio-lingual learning of Japanese as a second language. *Journal of Educational Research*, 56, 463-469.

Gu, P.Y. (1996). Robin Hood in SLA: What has the learning strategy researcher taught us? *Asian Journal of English Language Teaching*, 6, 1-29.

Hards, S.T. (1998). Erase English requirement. *The Japan Times Shukan ST*, 48 (35), 28.

Hatano, K. (1993). *Bungaku kyoiku wa naze hitsuyou ka?* (Why is the study of literature necessary?). In Takio Hida & Junya Noji (Eds.), *Kokugo kyoiku kihon ronbun shuusei* (A collection of foundational essays in Japanese education) (pp. 457-459), Vol. 16.

Hinds, J. (1983). Contrastive rhetoric: Japanese and English. *Text*, 3 (2), 183-195.

Hinds, J. (1990). Inductive, deductive, quasi-inductive: expository writing in Japanese, Korean, Chinese, and Thai. In Connors and Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 87-110). Alexandria, VA: TESOL.

Hino, N. (1988). *Yakudoku*: Japan's dominant tradition in foreign language learning. *JALT Journal*, 10 (1&2), 45-55.

Inoue, T. (1993). *Bungaku kyoiku ni okeru: Shudai* ("Topic" according to literature education). In Takio Hida & Junya Noji (Eds.), *Kokugo kyoiku kihon ronbun shuusei* (A collection of foundational essays in Japanese education) (pp. 464-468), Vol. 16.

Ishizuka, M. (1997). Japan needs to have a talk about English. *The Nikkei Weekly*, June 23, page 14.

Jannuzi, C. (1994). Team teaching the reading class. In M. Wada & T. Cominos (Eds.). *Studies in team teaching* (pp. 119-131). Tokyo: Kenkyusha.

Kawasaki, A. (1998). Survey of high school English teachers in Fukui Prefecture. In *Japanese and English rhetorical strategies: A contrastive analysis with pedagogical implications*. Unpublished graduation thesis: Fukui University.

Keizai doyukai: Jidai kangaerukai (Commission on the Economy: Focus Group on Thoughts for a New Generation). (1998). *Nihon no kyoiku wa doko ni mondai ga aru no ka?* (Where are the problems in the Japanese education system?). In Otake, M. (Ed.). *Nippon no yume to muda* (Japan's dreams and impossibilities) (pp. 202-221). Tokyo: Media Factory.

Kimura, S. & Visgatis, B. (1996). High school English textbooks and college entrance examinations: A comparison of reading passage difficulty. *JALT Journal*, 18 (1), 81-95.

Kitao, K. (1979). Difficulty of international communication—Between Americans and Japanese. *Doshisha Literature*, 29, 155-169.

Kitao, K., Broderick, V., Fujiwara, B., Kitao, S.K. & Sackett, L. (1985). *American Mosaic*. Tokyo: Eichosha Shinsha.

Kitao, K. & Kitao, S.K. (1989). *Intercultural communication: Between Japan and the United States*. Tokyo: Eichosha.

Kitao, K. & Kitao, S. K. (1995). *English teaching: Theory, research, practice*. Tokyo: Eichosha.

Kitao, K., Kitao, S.K. & Yoshida, H. (1985). *Daigakusai no eigo dokkai sokudo no kenkyu* (A study of college students' reading speed). *Chubu Chiku Eigo Kyoiku Gakkai Kiyo*, 14, 168-174.

Kitao, K. & Miyamoto, H. (1983). *Daigakusei no eigo dokkairyoku no mondaiten—gotou no keiko to suii* (Japanese college students' problems in reading English—tendencies and changes in errors). *Doshisha Studies in English*, 32, 118-147.

Kitao, K. & Yoshida, S. (1985). *Daigakusei no eigo dokkairyoku to sono speed no kenkyu* (A study of college students' English reading comprehension and speed). *Chubu Chiku Eigo Kyoiku Gakkai Kiyo*, 14, 28-34.

Kitao, K., Yoshida, S. & Yoshida, H. (1986). *Daigakusei no eigo dokkairyoku no mondaiten—gotou no ruikei to genin* (Causes of Japanese college students' problems in reading English). *Chubu Chiku Eigo Kyoiku Gakkai Kiyo*, 15, 8-13.

Kitao, S.K., Kitao, K., Nozawa, K. & Yamamoto, M. (1985). Teaching English in Japan. In: K. Kitao, K. Nozawa, Y. Oda, T. Robb, M. Sugimori, & M. Yamamoto (Eds.), *TEFL in Japan: JALT 10 Shunen Kinen Ronbunshu* (JALT 10th Anniversary Collected Papers), 127-138.

Kobayashi, H. (1984). Rhetorical patterns in English and Japanese. *TESOL Quarterly*, 18 (4), 737-738.

Kobayashi, Y. (1975). A new look at reading in the college program. *English Teaching Forum*, 13 (3-4), 188-195.

Kuroiwa, Y. (1997). Application and English fluency of minority language groups in the United States. *Gengo Kyoiku Kenkyu*, 8, 1-26.

Law, G. (1994). College entrance exams and team teaching in high school English classrooms. In M. Wada & T. Cominos (Eds.). *Studies in team teaching* (pp. 90-102). Tokyo: Kenkyusha.

Law, G. (1995). Ideologies of English language education in Japan. *JALT Journal*, 17 (2), 213-224.

Leonard, T.J. (1998). Japanese university entrance examinations: An interview with Dr. J.D. Brown. *The Language Teacher*, 22 (3), 25-27.

Matanle, P. (1996). History re-bleats itself. *Authentically English*, 3, 22-23.

McNabb, R. (1996). On eliminating English from university exams. *The Daily Yomiuri*, October 28, 12.

Mulvey, B. (1992). *Japanese and English rhetorical strategies: A contrastive analysis with pedagogical implications*. Unpublished M.A. thesis, California State

University, San Bernardino.

Mulvey, B. (1998). Entrance exams—the reading example. *ONCUE*, 6 (3), 5-12.

Nishijima, Hisao. (1995). *Nihonjin eigo no jakuten wa nani ka?* (What are the weak areas in Japanese learners of English?). *Eigo Kyoiku*, 44: 28-30.

Pai, A. (1996). Overcoming the barriers. *JET Journal*, summer, 151-157.

Ricento, T.K. (1987). *Aspects of coherence in English and Japanese expository prose*. Unpublished Ph.D. dissertation, University of California at Los Angeles.

Richards, J., Platt, J. & Weber, H. (1985). *Longman dictionary of applied linguistics*. Essex: Longman Group.

Robb, T.N. & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5 (2), 239-251.

Rohlen, T. (1983). *Japan's high schools*. Berkeley, CA: University of California Press.

Rumelhart, D.E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W.F. Brewer (Eds.). *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum.

Saeki, T. (1992). *Kagakutekina gaikokugo gakushuuhou* (Scientific methods for studying a foreign language). Tokyo: Kodansha.

Sakamoto, Takahiko. (1995). Problems in translation: A Japanese-to-English example. In T. Harris & R. Hodges (Eds.). *The literary dictionary* (p. 261). Newark, NJ: International Reading Association.

Sanford, A.J. & Garrod, S.C. (1981). *Understanding written language*. New York: John Wiley & Sons.

Shimaoka, T. & Yashiro, K. (1990). *Team teaching in English classrooms: An intercultural approach*. Tokyo: Kairyudo.

Sturman, P. (1989). Team teaching in Japan: the Koto-ku project. *JALT Journal*, 11 (1), 68-77.

Takahashi, M. & Takahashi, Y. (1984).*Dokkai* speed *no kijun wo doko ni okuka?* (How to set the criteria for reading speed?). *Eigo Kyoiku*, 33 (9), 36-39.

Takefuta, Y. (1982). *Nihonjin-eigo no kagaku: Sono genjo to asu e no tembo* (Scientific analysis of the English of the Japanese people). Tokyo: Kenkyusha.

Tanaka, C. (1985). *A Study of the Effectiveness of Reading Instruction at the College Level in Japan Based on Psycholinguistic Theory*. Unpublished Ph.D. dissertation, University of Kansas, DAI 8608453.

Tsukada, M. (1991). Student perspectives on juku, yobiko, and the examination system. In B. Finkelstein, A.E. Imamura, & J.J. Tobin (Eds.), *Transcending stereotypes: Discovering Japanese culture and education* (pp. 178-182). Yarmouth, ME: Intercultural Press.

Vanderford, S. (1997). Oral English in college entrance exams. *ONCUE*, 5 (3), 19-24.

Weaver, C. (1980). *Psycholinguistics and reading: From process to practice*. Cambridge, MA: Winthrop.

Yoshida, H. (1985). *CAI sokudoku kunren* (Speed reading: Training by computer). In K. Kitao, K. Nozawa, Y. Oda, T. Robb, M. Sugimori, & M. Yamamoto (Eds.), *TEFL in Japan: JALT 10 shunen kinen ronbunshu* (JALT 10th Anniversary Collected Papers) (pp. 45-53). Tokyo: JALT.

Yoshida, S. (1985). *Daigakusei no eigo dokkairyoku: Kodoku no shiryo toshite* (Japanese college students' reading ability: Data from college English courses). In K. Kitao, K. Nozawa, Y. Oda, T. Robb, M. Sugimori, & M. Yamamoto (Eds.), *TEFL in Japan: JALT 10 shunen kinen ronbunshu* (JALT 10th Anniversary Collected Papers) (pp. 117-125). Tokyo: JALT.

Yoshida, S. & Kitao, S.K. (1986). *Itsutsu no dokkai test wo riyoshita daigakusei no Eigo dokkai sokudo oyobi rikaido no kenkyu* (Japanese college students' English reading comprehension ability and speed—A study based on five tests). *Chubu Chiku Eigo Kyoiku Gakkai Kiyo*, 15, 183-188.

Yukawa, E. (1994). Team teaching and changes in teaching routines in a Japanese high school reading classroom. In M. Wada & T. Cominos (Eds.), *Studies in team teaching* (pp. 42-60). Tokyo: Kenkyusha.

Yutani, Y. (1977). Current English: Translation of news articles and "nonsequence" of tenses. *Academic Bulletin: Kyoto Bulletin of Foreign Studies*, 18, 52-63.

(Received April 22, 1998; revised October 29, 1998)

Reviews

The Language Instinct: How the Mind Creates Language. Steven Pinker. New York: Harper Perennial, 1994. 496 pp.

Reviewed by
Robert Blaisdell
Monterey Institute of International Studies

Since the fall of the behaviorist paradigm at the hands of Lenneberg, and Chomsky's irrefutable poverty-of-stimulus argument, innateness theories about the nature of human language have gained considerable ground. A great deal of theory and research has developed over the decades and the fires of debate around the innateness-versus-empiricism issue have burned at varying levels of intensity. Steven Pinker's voice rings out powerfully for the view that human beings are structurally designed by nature to develop and use one of our most definitive characteristics, language.

Pinker's *The Language Instinct* is a tour de force exposition on the nature of language. Arguing that language is an innate capacity of human beings, Pinker demonstrates through observation, reason, and theoretical research that language must be more deeply rooted than a mere set of behaviors which has accumulated through exposure to environmental input. Although his conclusions may side strongly with the innateness school, Pinker attempts to reconcile historical arguments by stating that even though language is encoded in the human chromosomes, it is nevertheless dependent on environmental stimuli to be triggered and patterned.

The book goes beyond a treatise on linguistics and selection theory. What adds to its force is that the medium is as much of the message as the content. Pinker's style is accessible, creative, contemporary, often contentious, and, above all, highly informed. He succeeds in bringing difficult arguments down from the ivory tower and making them available to the reader. Although this book is challenging, it delivers substantial rewards to those interested in languages, linguistics, and what the human brain and human language reveal about each other. Classroom pedagogues are left to themselves to apply the content of the book, but anyone interested in languages on any level will benefit from reading it.

Testing in Language Programs. James Dean Brown. Upper Saddle River, New Jersey: Prentice Hall Regents, 1996. 324 pp.

Reviewed by
Ian G. Gleadall
Tohoku Bunka Gakuen University

Books on testing generally fall into two categories: those dealing with the practical aspects of constructing and evaluating tests and those reviewing theories of test construction and development. *Brown's Testing in Language Programs* (TILP) is a new departure, providing comprehensive coverage of the theory but also going deeply into the appropriate usage of many of the statistical functions commonly used in evaluating language tests (see also Brown, 1989). The text is generally very clear and easy to read, especially with its unusually large typeface, but the section on measuring and displaying data contains some errors which (evidently repeated from a pedigree of other EFL texts) are particularly cause for concern in such a basic book.

TILP's nine chapters begin with an overview of the content and end with a summary, often in list form, followed by consolidation questions and application exercises. The Table of Contents presents only the chapter titles, whereas the inclusion of subheadings would have been useful given Brown's central theme of criterion-referenced testing (CRT) versus norm-referenced testing (NRT) and the consequent subdivision of most chapters into these sections.

The NRT versus CRT organizational approach to testing has obvious advantages in dealing with the statistical analyses of different types of tests, but Brown's discussion of the properties of these two categories might be considered too simple. For example, other classifications (e.g., subjective versus objective; long versus short) are included in the debate as if they have the same demarcation as CRTs and NRTs, which they do not. Brown (p. 8) also tries to fit the four primary language testing functions into the CRT/NRT scheme, claiming that they "correspond neatly" with NRTs (for proficiency and placement decisions), and CRTs (for achievement and diagnostic decisions). His separation of CRTs and NRTs involves acceptance of the assertion that CRTs measure "specific, objectives-based language points," while NRTs measure vaguely defined "general language abilities or proficiencies" (see Table 1.1 on p.3). However, Cartier (1968) has characterized NRTs as testing a sample of the course objectives, while CRTs ideally should test all the objectives (hence the 'subjective' versus 'objective' comparison, for example, is inappropriate); and Brown's contention that NRTs are "long" and CRTs "short" is just the opposite of what Cartier (1968) claimed.

The first half of Chapter 2 (pp. 21-35) introduces the major theoretical and practical issues in testing and is well written in a series of short, concise sections. Theoretical issues include language teaching methodology, skills, competence and performance, and discrete point versus integrative testing. These are followed by two useful checklists for evaluating testing programs. However, the lack of examples of (or even parts of) actual tests is a missed opportunity to consolidate the characteristics of CRTs and NRTs.

Chapter 3 deals with developing and improving test items, with checklists summarizing the guidelines for most item formats and an analytic scale for rating composition tasks. The application exercises at the end of this chapter are very useful and working through them will provide a firm grounding in what this chapter has to teach about item analysis. However, some small inconsistencies in the usage of terms could confuse the neophyte: "correct answer" and "key" are both used, with no mention that they mean the same thing; similarly with "miskey" (which presumably means a distractor, not the key, that was chosen by the testee) and "missed the item" (p. 79).

Chapters 4 and 5 cover the arithmetical concepts required to understand the topics of correlation, validity, and reliability covered in Chapters 6-8. Chapter 4 deals with counting and measuring, presentation of statistical data in tabular form, displaying data, and central tendencies. Chapter 5 ("Interpreting test scores") uses probability to introduce the normal distribution, and presents a concise explanation of standard scores, including *z*, *T* and CEEB (as used, for example, to report TOEFL scores). However, using stars and crosses to illustrate bar charts and histograms is confusing and unnecessary in this age of computer-aided chart construction. More important, though, is the failure to clearly distinguish 'continuous' from 'discontinuous' data, and consequently to distinguish histograms from bar charts (e.g. Fig. 5.1, p. 125): errors that require urgent correction. It is also inappropriate to use the number of languages a person speaks to illustrate a "ratio scale" (pp. 97-98), since it has an absolute zero but no one speaks "zero" languages; or to use decimal places merely for neatness (Table 4.7, p. 111), for example where "N" is the number of students who took a given test (integers/students cannot be divided into hundredths, which is what two decimal places implies).

Chapter 6 is very lucid, particularly the section on correlation coefficients for random numbers, and the discussion of the importance of considering the relative magnitude of the correlation coefficient in different situations. Brown's discussions of reliability (Chapter 7) and validity (Chapter 8) are also clear and thorough. However, ANOVA and omega squared analyses (Tables 8.2 and 8.3) are tantalizingly mentioned while stating that they are "beyond the scope of this book" (p. 242). Brown should have omitted them, or explained them fully.

The final chapter places testing as a central issue in curriculum planning. This is followed by the key to the application questions. However, I was frustrated not to find answers to some of the review questions (such as that on p. 147, asking the reader to calculate probabilities). The final reference section is an extensive bibliography. There is neither glossary nor appendices (e.g. statistical tables, formulae, or examples of test formats).

There are some surprising omissions from TILP: The words "computer" and "software" appear only on pp. 42 and 91. In a text of this nature, one would expect some discussion of statistics software packages, or at least a mention of spreadsheets, and also a list of suitable software products and references for their use by the digitally challenged. The communicative paradigm is only briefly mentioned by Brown, who could have been more informative about recent developments. Most surprising of all, however, I could find no mention of the important concept of washback in TILP (cf. Brown, 1997). Communicative testing and washback are important current issues in language testing and should be included. There is also no discussion of the meaning and fundamental importance of objectives in the construction of both syllabuses and tests, despite the inclusion of terms such as "course objectives" (p. 14), "specific instructional objectives" (p. 15), and the subheading "goals and objectives" (p. 272). In a text emphasizing the reliance of CRTs on the effective stating of objectives, I would expect to see a brief section on the writing of behavioral objectives or at least some references to guide the reader.

To summarize, TILP provides a readable approach to statistics as used in language testing and deals thoroughly with the practical, technical aspects of test evaluation that should be addressed by those responsible for assessment in and evaluation of language programs. However, attention to the omissions and small errors is required in a revised second edition, with the detailed arithmetic perhaps moved to appendices. Otherwise, my only hesitation in recommending this very useful book is its over-simplistic division between CRTs and NRTs.

References

Brown, J.D. (1989). *Understanding research in second language learning*. Cambridge: Cambridge University Press.

Brown, J.D. (1997). The washback effect of language tests. *University of Hawaii working papers in English as a second language*, 16 (1), 27-46.

Cartier, F. (1968). Criterion-referenced testing of language skills. *TESOL Quarterly*, 2, 27-32.

Using Corpora for Language Research. Jenny Thomas and Mick Short (Eds.). London: Longman Group Limited, 1996. 301 pp.

Reviewed by
Jim Ronald
Hiroshima Chapter

Using Corpora for Language Research (UCLR) is a collection of sixteen papers relating to the use of language corpora (computer-based collections of written and/or spoken texts) in various kinds of language research. The papers are divided into four sections: an introductory section focusing on the importance of corpora in language research; a section on various corpus-based language studies; a section about technology-related applications of research using corpora; and a final section, perhaps of most direct relevance to language teachers, entitled "Wider Applications of Corpus-based Research."

UCLR claims to be for people who are interested in language work but who are not corpus specialists. As far as possible, I will consider this book from this non-specialist perspective by asking some general questions.

First, does the collection address basic theoretical and practical questions about using a corpus for language study? Related questions are "Why bother with a corpus? Isn't my intuition enough?" or "How, practically, can corpus work affect what a language teacher does?" or "How big should a corpus be?" Most of these issues are addressed, or acknowledged here, although they are not always easy to find. Sampson's paper (Chapter 2) provides a "road to Damascus" account of his conversion to corpus linguistics, from a generative grammar background in which examples of real language count for very little. He was persuaded of the value of corpus work by the undeniable evidence of the widespread, if still rare, use of a linguistic feature (central embedding) that theorists had intuitively decided should not exist. For those not from such a background, and perhaps more easily convinced of the value of corpus work, Alderson very simply states what a corpus offers: "Linguists can now have recourse, not just to their intuitions, but also to others' language use" (p. 248).

This brings us to the next question: "How, practically, can corpus work affect what a language teacher does?" The articles by Mindt on corpus linguistics and the foreign language teaching syllabus (Chapter 14) and Alderson on the possible uses of corpora in language testing (Chapter 15) together provide a good introduction to many of the theoretical and practical considerations relating to teaching applications of corpus work. Mindt, for example, compares the ordering and presentation of future time orientation, modals, and conditional in English textbooks in Germany with their relative frequency and typical use as measured using corpora of spoken English. He concludes that there is evidence justifying a number of changes in the textbooks' treatment and ordering of these structures. It should be noted that such research could not have been done before computers and software made the analysis of sufficient volumes of language possible, thereby producing reliable measurements of frequency and the typical use of aspects of general language.

Alderson (Chapter 15) speculates as to how corpora could be used in language assessment. He suggests possible applications of corpora, such as using them as a source of real texts in testing, identifying frequent lexical items for use in texts, or using a corpus of learners' texts to identify problem areas of language. It is surprising, however, that Alderson's paper is wholly speculative and that he should not have encountered actual instances of corpora being used in language assessment. The writer of this review is surely not alone in using a corpus or real examples from corpus-based resources in the testing of grammatical structures and lexical items.

"How big should a corpus be?" is a more complex question than it might seem, as this depends on the purpose of the corpus, what texts the corpus should comprise, and, if a corpus is composed of more than one type of language (e.g., American spoken, British written, newspapers), what proportions of each type should be included. For some purposes, most prominently computational lexicography, corpora of between 100 million and 300 million words are not unusual and are necessary to enable an accurate description of the typical use of less common syntactically variable lexical items. This issue is touched on by Della Summers of Longman Dictionaries (Chapter 16), but is somewhat slanted by the commercial orientation of her paper.

Elsewhere in this text, research is reported using surprisingly small corpora. For example, in one paper (Chapter 6), subcorpora as small as 8,000 words and comprising only four or five texts, such as letters or academic papers, are used to provide general statements about language use in that type of text. However, individual writing styles and topic choice are such that observations about language based on such small corpora cannot reliably be used to make generalizations about typical language use. While there is, undoubtedly, a case for smaller corpora (e.g., in ESP), the issue is not considered here at all.

With its wide range of topics, this collection appears initially to be providing an overview of the current state of corpus-based language research, or even to be demonstrating the truth of the first sentence in the book, that "Corpus linguistics has now become mainstream" (p. ix).

If this is its aim, it falls short of achieving it in a couple of important respects. This collection of articles has been assembled in honor of Geoffrey Leech, a central figure in corpus linguistics ever since this mainstream was just a trickle. Whatever the intentions of the editors, however, this book is not a demonstration of the "mainstreamness" of corpus linguistics, nor of Leech's wide-reaching influence in this expanding field, as we might expect such a *festschrift* to be. Rather, it appears more as a claim by Lancaster University for preeminence in this area. This is evident, among other things, in the large proportion of articles here written by Lancaster University faculty and in the virtual exclusion of other important centers of corpus work. In addition, most of the studies reported in this volume are major projects by important figures in linguistics undertaken with funding from government or industry, and using very large corpora or involving detailed manual tagging. Although figures are not available, I would imagine that the majority of corpus-related research projects around the world are smaller, using fairly simple concordancing programs such as Johns & Scott's *MicroConcord* (1993) with untagged corpora of tens or hundreds of thousands of words rather than tens or hundreds of millions, or using the resources of a publicly available (at a price) corpus such as COBUILD's Bank of English. Including one or two accounts of smaller projects would have been helpful to those who are not specialists in the field.

For someone new to corpus linguistics the above weaknesses may not be too apparent. Their consequences, however, could be that the reader gains a distorted and incomplete picture of the world of corpus linguistics, perhaps being left with the impression that corpus linguistics is largely restricted to a small group of researchers based in one British university, or feeling that the means to undertake language research using corpora are beyond their reach. This would be unfortunate as neither impression would be correct. Corpus work is increasingly popular in many countries around the world, including Japan, and part of its appeal is that, both technically and financially, it is relatively accessible.

In terms of providing an introduction to corpus linguistics, there are a few papers in *Using Corpora for Language Research* that do address many fundamental issues relating to corpus work. As a whole, though, I would feel bound to recommend other texts to a colleague interested in knowing something about corpus linguistics. Aijimer & Altenberg's *English Corpus Linguistics* (1991) provides a more rounded and accessible introduction to the subject. For those interested in actually developing and using their own corpora, and in classroom applications of corpus work, Wichmann, Fligelstone, McEnery & Knowles's *Teaching and Language Corpora* (1997) is a good place to start.

References

Aijimer, K. & Altenberg, B. (Eds.). (1991). *English corpus linguistics*. London: Longman.
Johns, T. & Scott, M. (1993). *MicroConcord*. Oxford: Oxford University Press.
Wichmann, A., Fligelstone, S., McEnery, T. & Knowles, G. (Eds.). (1997). *Teaching and language corpora*. London: Longman.

Teacher Cognition in Language Teaching: Beliefs, Decision-Making and Classroom Practice. **Devon Woods. Cambridge: Cambridge University Press, 1996. 316 pp.**

Reviewed by
Kazuyoshi Sato, Nagoya University &
Tim Murphey, Nanzan University

At the 1997 JALT Conference Devon Woods asked, "What do we mean when we say 'teaching'?" His talk was based on research reported in *Teacher Cognition in Language Teaching* (TCLT), a work which examines the relationship between teachers' beliefs and their practices.

In foreign language teaching the significance of research on teachers' beliefs with regard to practices has been only recently recognized, and little is known in general about how teachers make sense of teaching and how they actually teach in the classroom. Kleinsasser and Savignon (1991) claim that "little systematic inquiry has been conducted into language teacher perceptions and practices" (p. 291). TCLT addresses this lacuna by looking at three broad areas: (1) The teaching structures of eight ESL teachers; (2) their planning procedures; and (3) their interpretive processes.

TCLT is made up of 10 chapters. Chapter 1 presents a rationale for studying the teachers he chooses and identifies three research questions. Chapter 2 discusses the research methodology, which employs triangulation or multiple data sources such as ethnographic interviews, logs, video-based recall, and documents such as lesson plans. Woods derives his particular method from ethnography and cognitive studies. Chapters 3 and 4 examine the structure of teaching and review models of teachers' decision-making, which represent the cycle of planning, action, and interpretation. Chapter 5 delineates the planning process of teachers and presents a new dynamic model which includes both lower and higher levels of planning and decision-making. Chapter 6 uncovers teachers' decision-making or interpretive processes and emphasizes the role of experienced structures, which are related to teachers' beliefs. Chapter 7 presents an integrated view of the network of beliefs, assumptions, and knowledge (BAK) which teachers hold, and concludes that teachers structure their teaching depending on their BAK. Woods offers an in-depth analysis of one teacher's language learning and teaching experiences in order to exemplify the development of a BAK. He concludes that, "BAK develops through a teacher's experiences as a learner and a teacher, evolving in the face of conflicts and inconsistencies." (p. 212). Chapter 8 examines the influence of BAK on teachers' practices, curricula, and theory. The author claims that the pervasiveness of BAK influences "the teachers' organization of thoughts, decisions, and aspects of the course" (p. 249), indicating the strong relationship between beliefs and practices. Chapters 9 and 10 elaborate on teacher change and curricular evolution.

The strength of TCLT lies in the scrutiny of teachers' beliefs in relation to their practices, focusing on events, planning, and decision-making processes. In particular, Woods reveals the strong effect of previous teaching experiences on a teacher's BAK. He affirms that, "Teachers seemed to prefer and trust experienced structures and tended to avoid structures that were completely new to them" (p. 182). The importance of actual teaching experiences implies a need to reconfigure the traditional knowledge-transmission model of teacher education. The author proposes a "different way of thinking about teaching" (p. 297) in contrast to the research-driven top-down change. He claims that "teacher change can be encouraged but not mandated" (p. 293).

One weakness of TCLT lies in the scant empirical evidence attesting actual teacher change or development. The author acknowledges that seven teachers out of eight did not show any clear change. He attributes the lack of evidence of change to "the developing skill of the interviewers" and "the willingness of the subject to delve into background experiences" (p. 203). Are we to conclude, therefore, that beliefs formed by previous experience cannot be changed? Even in the case of teacher B, described as the 'best example,' L2 learning experiences and past teaching experiences influenced his beliefs, but there was no change reported in his beliefs during this study. Moreover, readers might wonder how new teaching experiences affect BAK. The author suggests that "teachers are in constant change" (p. 257), if they are offered "opportunities for reflection and interaction as a catalyst for change" (p. 297). While we intuitively agree with the conclusion, we did not see much supportive evidence in this study.

In addition to that, we feel that Woods has overemphasized internal processes and disregards the impact of external contexts that can help create and foster experimentation and internal changes. He maintains that, "Because this study is a study of individual cognitions and not of social conventions, this is an empirical question I have not attempted to answer" (p. 115). Nevertheless, in his analysis, he refers to external contexts as significant factors several times, finally acknowledging that both internal and external elements are necessary for the change to occur. He suggests that internal elements include a teacher's "interest in change" and "conceptual readiness for change" (p. 294). The external elements are the teaching culture or social environments where teachers interact with other teachers, share views, ideas and materials, and have opportunities to experiment.

He finally concludes that, "Reflective teaching develops out of social environments in which experimentation . . . appear natural" (p. 298). This conclusion is a big leap from his original stance which did not include contexts. He notes (p. 297) that the teachers who did not report change might have felt isolated or been in less collaborative cultures, which are often the most common teaching cultures. In fact, some researchers point directly to the significance of institutional development for fostering an environment for teacher development (Fullan, 1991; Lieberman & Miller, 1990). Future research needs to clarify how teachers' beliefs and practices can develop within certain teaching cultures or contexts and how these environments can be structured.

Despite these weaknesses, Woods does clarify the complexity of teachers' decision-making processes in connection with their pervasive BAK. In particular, he stresses the significance of teaching experiences. Thus, TCLT encourages teachers to try new ideas, interact with other teachers, share ideas and materials, and develop curricula collaboratively, thereby creating supportive contexts for themselves and others. The shift from a 'static' view of top-down teacher education to one of 'dynamic' teacher development and curricular development involving the use of a teacher's evolving network of beliefs, assumptions, and knowledge is one we hope that more teacher trainers and teachers will make. This organic evolution is a result of "experiences that resulted in a conflict with the BAK's current state" (p. 248), and creating safe, collaborative environments for such experiences needs much more of our attention.

References

Fullan, M.G. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Kleinsasser, R.C. & Savignon, S.J. (1991). Linguistics, language pedagogy, and teachers' technical cultures. In J. E. Alatis (Ed.). *Georgetown University Round Table on Language and Linguistics 1991* (pp. 289-301). Washington, DC: Georgetown University Press.

Lieberman, A. & Miller, L. (1990). Teacher development in professional practice schools. *Teacher College Record*, 92 (1), 105-122.

Information for Contributors

All submissions must conform to *JALT Journal* Editorial Policy and Guidelines.

Editorial Policy

JALT Journal, the refereed research journal of the Japan Association for Language Teaching (*Zenkoku Gogaku Kyoiku Gakkai*), invites practical and theoretical articles and research reports on second/foreign language teaching and learning in Japanese, Asian, and other international contexts. Areas of particular interest are:

1. curriculum design and teaching methods
2. classroom-centered research
3. cross-cultural studies
4. testing and evaluation
5. teacher training
6. language learning and acquisition
7. overviews of research and practice in related fields

The editors encourage submissions in five categories: (1) full-length articles, (2) short research reports (*Research Forum*), (3) essays on language education or reports of pedagogical techniques which are framed in theory and supported by descriptive or empirical data (*Perspectives*), (4) book and media reviews (*Reviews*), and (5) comments on previously published *JALT Journal* articles (*Point to Point*). Occasionally *JALT Journal* will issue a Call for Papers for theme-based issues. Articles should be written for a general audience of language educators; therefore statistical techniques and specialized terms must be clearly explained.

Guidelines

Style

JALT Journal follows the ***Publication Manual of the American Psychological Association***, 4th edition (available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20784, USA). Consult recent copies of ***JALT Journal*** or ***TESOL Quarterly*** for examples of documentation and references.

Format

Full-length articles must not be more than 20 pages in length (6,000 words), including references, notes, tables and figures. ***Research Forum*** submissions should not be more than 10 pages in length. ***Perspectives*** submissions should not be more than 15 pages in length. ***Point to Point*** comments on previously published articles should not be more than 675 words in length, and ***Reviews*** should generally not be longer than 500-750 words. All submissions must be typed and double-spaced on A4 or 8.5"x11" paper. The author's name and identifying references should appear only on the cover sheet. Authors are responsible for the accuracy of references and reference citations.

Materials to be submitted

1. Three (3) copies of the manuscript, with no reference to the author. Do not use running heads.
2. Cover sheet with the title and the author name(s)
3. Contact information, including the author's full address and, where available, a fax number and electronic mail address
4. Abstract (no more than 150 words)
5. Japanese translation of the title and abstract, if possible (less than 400 *ji*)
6. Biographical sketch(es) (no more than 25 words each)
7. Authors of accepted manuscripts must supply camera-ready copies of any diagrams or figures and a disk copy of the manuscript (RTF or ASCII)

Evaluation procedures

All manuscripts are first reviewed by the editorial board to insure they comply with *JALT Journal* Guidelines. Those considered for publication are subject to blind review by at least two readers, with special attention given to: (1) compliance with *JALT Journal* Editorial

Policy, (2) the significance and originality of the submission, and (3) the use of appropriate research design and methodology. Evaluation is usually completed within three months.

Restrictions

Papers submitted to *JALT Journal* must not have been previously published, nor should they be under consideration for publication elsewhere. *JALT Journal* has First World Publication Rights, as defined by International Copyright Conventions, for all manuscripts published. We regret that manuscripts or computer disks cannot be returned. In the interests of facilitating clarity, the editors reserve the right to make editorial changes to accepted manuscripts.

Full-Length Submissions, Research Forum, and Point to Point Submissions

Please send submissions in these categories or general inquiries to:

Sandra Fotos, Editor
School of Economics, Senshu University,
2-1-1 Higashi Mita, Tama-ku, Kawasaki,
Kanagawa-ken 214-0033, Japan

Perspectives

Please send submissions in this category to:

Nicholas O. Jungheim, Associate Editor
Faculty of Law, Aoyama Gakuin University
4-4-25 Shibuya, Shibuya-ku, Tokyo 150-0002, Japan

Reviews

The editors invite reviews of books, tests, teaching systems, and other publications in the field of language education. A list of publications which have been sent to JALT for review is published monthly in *The Language Teacher*. Please send submissions, queries, or requests for books, materials and review guidelines to:

Patrick Rosenkjar, Book Reviews Editor
Temple University Japan
2-8-12 Minami Azabu
Minato-ku, Tokyo 106-0047, Japan

Japanese-Language Manuscripts

JALT Journal welcomes Japanese-language manuscripts on second/foreign language teaching and learning. Submissions must conform to the Editorial Policy and Guidelines given above. Authors must provide a detailed abstract in English, 500-750 words in length. Refer to the Japanese-language Guidelines for details. Please send Japanese-language manuscripts to:

Shinji Kimura, Japanese-Language Editor
Faculty of Law, Kwansei Gakuin University,
1-1-155, Uegahara, Nishinomiya, Hyogo 662-0886, Japan

Address for Inquiries about Subscriptions or Advertising

JALT Central Office
Urban Edge Building 5F
1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel.: 03-3837-1630; Fax: 03-3837-1631
(From overseas: Tel.: 81-3-3837-1630; Fax: 81-3-3837-1631)

JALT Journal 日本語論文投稿要領

JALT Journalでは、日本語で執筆された第二言語・外国語教育についての論文を募集いたします。第二言語教育／学習研究の文脈を踏まえ、しっかりした研究計画に基づいている、実践への応用の可能性を示した理論的・実証的研究や理論的裏づけを持った実践報告などで、教育、教授法、応用言語学などの今日的問題を扱ったものを歓迎します。

文体は一般的な学術論文のスタイルを使ってください。章だてのしかたや参考文献のデータの書き方などは、Publication Manual of the American Psychological Association (4th. ed.) の定める方式にできるだけ近い形にしてください。ご不明の場合は、JALT Journal の英語論文を参考にするか、日本語編集者までお問い合わせください。また、JALT Journal の読者は現場の教師が主ですから、特殊な専門用語や統計的手法は、わかりやすいように定義したり説明したりしてください。

原稿の長さは、参考文献リストも含め400字づめ原稿用紙30枚以内です。A4の用紙に横書きで、1行40字、1ページ30行で印刷してください。手書きの原稿は受け付けられません。図表を入れる場合は、JALT Journal のページのサイズに合わせて作成してください(縮小コピー可)。執筆者の名前や所属機関名は、原稿には書かないでください。ご提出いただくものは、以下の通りです。

- 原稿3部
- 執筆者の名前と所属機関名を書いた表紙
- 400字以内の和文要旨
- 英文のタイトルと、500～750語の英文要旨
  英文要旨が用意できない場合は、400字づめ原稿用紙3枚程度の詳細な和文要旨をご提出ください。
- 100字以内の執筆者略歴
- 可能ならば、Macintosh のテキスト形式で保存したファイル

査読の方法は、編集委員会で投稿要領にあっているかどうかを確認したあと、少なくとも二人の査読者が査読を行います。査読者には執筆者の名前は知らされません。査読の過程では特に、原稿がJALT Journal の目的にかなっているか、言語教育にとって意味があるか、独創性はあるか、研究計画と方法論が適切かが評価されます。査読は通常二か月以内に終了します。

JALT Journal に投稿する原稿は、すでに出版されているものや他の学術雑誌に投稿中のものは避けてください。JALT Journal は、そこに掲載されるすべての論文に関して国際著作権協定による世界初版権をもちます。

お送りいただいた原稿は返却できません。

投稿原稿の送り先とお問い合わせは以下にお願いします。
〒620-0886　兵庫県西宮市上ヶ原1-1-155関西学院大学法学部
JALT Journal 日本語編集者　木村真治

JALT Journal 第21巻　第1号
1999年4月20日　印刷
1999年5月1日　発行

編集人　サンドラ・フォトス
発行人　ジーン・ヴァン・トロイヤー
発行所　全国語学教育学会事務局
〒110-0016 東京都台東区台東1-37-9 アーバンエッジビル5F
TEL (03)3837-1630; FAX (03)3837-1631
印刷所　コーシンシャ株式会社
〒530-0043 大阪市北区天満2-13-3 TEL (06)351-8795

昭和五十四年四月二十一日 第三種郵便物認可
THE LANGUAGE TEACHER 特別増刊号

25th Anniversary

JALT 99

Japan Association for Language Teaching

全国語学教育学会

理論と実践
研究と教室の架け橋

Teacher Belief, Teacher Action:

Connecting Research and the Classroom

October 8 –11, 1999
Maebashi Green Dome, Gunma Prefecture
(about one hour from Tokyo)

1999年10月8日～11日
群馬県前橋市グリーンドーム
(東京より約１時間)

The 25th Annual International Conference on Language Teaching and Learning & Educational Materials Expo

第25回JALT年次国際大会語学教育/学習 & 教材展

Plenary Speakers
Dr. Richard Allwright, University of Lancaster, UK
Dr. Anna Uhl Chamot, George Washington University, USA
Dr. Elizabeth Gatbonton, Concordia University, Canada
Mario Rinvolucri, Pilgrims Ltd., UK

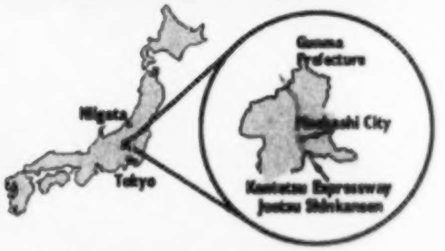
後援：群馬県教育委員会、前橋市、前橋市教育委員会、上毛新聞社、群馬テレビ、エフエム群馬

Endorsed by: Gunma Prefecture Board of Education, Maebashi City, Maebashi Municipal Board of Education, The Jomo Shimbun, Gunma TV, and FM Gunma

http://www.jalt.org/conferences/

Making Connections
Creating Communities
Deepening Cooperation

Maebashi Green Dome



For information or registration:
Contact JALT Central Office
Tel: 03-3837-1630
Fax: 03-3837-1631